Student Response in Programmed Instruction (1961)

Pages
570

Size
7 x 10

ISBN
0309359562

A. A. Lumsdaine, Editor; National Research Council

STUDENT RESPONSE IN PROGRAMMED INSTRUCTION

A Symposium on Experimental Studies of Cue and Response Factors in Group and Individual Learning from Instructional Media

Edited by A. A. Lumsdaine

Sponsored by
HEADQUARTERS
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE

Publication 943
NATIONAL ACADEMY OF SCIENCES—NATIONAL RESEARCH COUNCIL
Washington, D.C.
1961

Library of Congress Catalog Card Number 61-60098

PREFACE

The primary purpose in preparing the present volume has been to bring together and make available a group of studies which have considerable relevance both for programmed individual self-instruction by teaching machines and related devices, and also for sequenced, reproducible instructional programs presented by films and other media. Some of the distinctions and relationships between the specialized and more general forms of programmed instruction are discussed in Chapter 1.

The papers that have been collected and edited for this volume are concerned with the experimental study of procedures for guiding the implicit or overt responses of learners in order to increase the effectiveness of instruction. The present volume is in no sense a comprehensive text; rather, it is a symposium reporting the findings of relevant studies that developed out of military research programs and that would, in the main, have become increasingly inaccessible unless collected in such a volume.

All these studies involve conscious attention to the manipulation of appropriate implicit or overt responses by the learner, and most of them involve techniques for eliciting and/or guiding overt responses during the course of instruction. However, the studies do not represent a fully integrated programmatic effort. They were conducted at a number of locations over a considerable period of time and, whether performed on an in-service basis or under contract, reflected a policy of allowing considerable latitude to individual investigators in the formulation of problems and research designs. At the same time, the common concern with student response and methods for its effective control during instruction gives at least a quasi-programmatic and somewhat integrated character to this group of studies, as contrasted with symposia covering wider or more diverse ranges of experimental inquiry into methods of training and education. To relate the individual papers more closely in terms of a common conceptual framework, it was originally planned that they would be presented and discussed, prior to their final revision, at a meeting attended by the primary authors. The holding of such a meeting was precluded, however, by budgetary limitations, and the preparation and revision of the papers was accordingly handled by correspondence with the editor.

With two or three exceptions, these papers represent products of a series of studies conducted under the editor's direction, on a joint in-service and contractual basis, for the Research and Development Command of the U. S. Air Force during the period 1950-1957. The work reported in this volume was part of a wider group of studies originally conceived within the framework of research on instructional films, but later integrated with concurrent research on other forms of instructional media and training devices. A number of additional studies conducted in the context of the research program on instructional films have been omitted here, in order to keep the scope of the book within reasonable bounds and to permit integration of the work reported in terms of factors related to the manipulation of student response.

Among the studies excluded is a group of experiments concerned with basic factors in pictorial perception conducted under a contract at Cornell University; most of these have been subsequently published by J. J. Gibson, Olin W. Smith, and collaborators. Also excluded are a number of miscellaneous studies that do not fit within the basic framework of the present volume. The latter included both practical development efforts, such as the development of a semi-automated language instruction course (cf. Lumsdaine, 1959a), and experimental studies employing various techniques of observation and concerned with the manipulation of other classes of variables related to instructional effectiveness of films. Some of the latter studies were never completed because of programmatic changes within the cognizant Air Force agencies. Several of the completed studies are abstracted in the appendix to this volume. Since the principal concern here is with the implications of research for the development of a science of instructional programming, purely technological applications of research are not dealt with. For example, except for some brief summaries in the Appendix, none of the papers included here deal with the utilization of student-response data for product improvement through "feedback" to the designer of a specific instructional instrument. Also not specifically dealt with here are implications of some of the research for the design of hardware, either in the form of teaching machines or of devices for informational input to work-performance situations (cf. Hoehn and Lumsdaine, 1958). Instrumentation for the effective use of filmed demonstration, of the kind with which all the studies in Part I and several in Part II are concerned, has been briefly described elsewhere by the writer (Lumsdaine, 1959c).

The research summaries collected in the Appendix represent, in the main, studies that are of rather direct relevance to the central concern of the volume, but which, for one reason or another, could not be included in the main body of the text. In general the studies presented in the text of the volume, as well as those summarized in the Appendix, have hitherto either been unpublished or published only as mimeographed or multilithed reports. Most of these never received very wide circulation, and are now generally inaccessible. A few exceptions, in instances in which studies have been published in journals or other more widely available sources, are noted in the footnotes to individual chapter titles.

A number of abbreviations are used in identifying organizational components involved in the sponsorship of these studies; where these are not self-explanatory, the reader can identify them by reference to the introduction to the Appendix in the source book, "Teaching Machines and Programmed Learning," edited by A. A. Lumsdaine and Robert Glaser for the National Education Association (1960). This source also gives further information about the availability of various categories of Air Force reports, such as Technical Notes and Technical Memoranda, pertaining to earlier reports on which some of the present papers were primarily based. As in the Lumsdaine-Glaser source book (abbreviated TM&PL below), it has seemed advisable to include references to these sources, despite their limited availability, in order to identify the origins of the present papers.

The feasibility of ready reference to TM&PL is taken for granted on the assumption that it will be available to virtually all individuals actively interested in the present field of inquiry. It should also be noted that rather detailed abstracts of a number of the further references cited by various authors of papers in the present volume are contained in Appendix I of TM&PL. In the reference list for the present volume, papers abstracted or reprinted in full in TM&PL have been so identified, in addition to original source citation, to facilitate locating them in that reference volume. Relevant studies abstracted in TM&PL have not been abstracted in this volume, to avoid unnecessary duplication, but fuller reports of several of the papers abstracted in TM&PL are presented in the present volume.

As in TM&PL, all reference citations in this book are brought together in a single comprehensive bibliography, which also serves as an author index through the inclusion, for each entry, of the page numbers where reference to each entry is made. Brief annotations accompany a number of the references in the bibliography where it seemed that this would help to clarify the status of the paper cited, including its relation to earlier or later versions of the same report or related reports.

The editor would like to thank the several contributors for the work done in preparing papers, and for their tolerance in permitting editorial liberties with their work. He has made occasional minor stylistic changes in most of the papers, and in a few instances more extensive editorial changes. Where special additions or editorial notes have been added, these have generally been set off in brackets to identify them as such. Specific sponsorship of and participation in each of the studies is indicated in an initial footnote at the beginning of each paper.

The program of research on films from which the earlier of the studies here reported were derived was initiated under the editor's direction at the former Human Resources Research Laboratories (later Human Factors Operations Research Laboratories) of the Air Force, and the editor would like to express his appreciation to Dr. Karl D. Kryter, then director of HRRL, for administrative and policy support of this program. A number of the studies initiated under this aegis were completed following the transfer of the instructional film research activity to the Training Aids Research Laboratory of the Air Force Personnel and Training Research Center (AFPTRC). These included both in-service and contract studies, and were under the general direction of the editor in his capacity as Technical Director of the Training Aids Research Laboratory. A number of the studies were under the more specific monitorship of Dr. Sol M. Roshal.

The editor would also like to express his appreciation for support provided by Dr. Arthur W. Melton, Dr. Charles W. Bray, and other members of the Headquarters Staff of AFPTRC, and by Paul N. Smith, Henry L. Adams, and other staff personnel of the Training Aids Research Laboratory who assisted in administrative and technical matters connected with these studies. In the last group of studies to be completed (those comprising Part I of the present volume), extensive technical and administrative monitoring duties were very capably handled by Dr. Arthur J. Hoehn following the integration of the Training Aids Research

Laboratory with the AFPTRC Maintenance Laboratory at Lowry Air Force Base, Colorado. The editor would particularly like to thank Dr. Glen Finch of the National Academy of Sciences—National Research Council, both for the support he provided as a member of the Technical Advisory Staff at U. S. Air Force Headquarters during the earlier phases of research here reported, and for his later assistance and encouragement in making it possible to publish the present volume.

The physical work of editing the papers and of preparing the Appendix, bibliography, and introductory and concluding chapters was accomplished at the School of Education of the University of California, Los Angeles, and the editor would like to thank Howard E. Wilson, Dean of the School of Education, and his staff, for encouragement and assistance during the completion of the volume. The editor's thanks are expressed to Yasuko M. Filby for her assistance in preparing the research summaries in the Appendix, and to Donn L. Cochran, who assisted in the later phases of some of the editorial work. Valuable assistance in both editorial and administrative matters was also furnished by Nancy Hamilton. For their help in preparing the manuscript the editor would especially like to thank Darlene Lloyd and Nina Van Zoeren for their unstinting efforts, often involving long hours of work outside the normal daily schedule, without which the completion of the volume would have been impossible. The editor's special thanks are also due to Robert Hume for final editing of the completed manuscript of the Academy—Research Council and supervision of arrangements for preparation of proof. Finally, he would like to acknowledge his indebtedness to many individuals who, through their direct and indirect participation in the program of Air Force research under which these studies were completed, contributed to the formulation of research conceptualization and design. In this respect he would particularly like to acknowledge the contributions made by Fred D. Sheffield and Sol M. Roshal.

A. A. Lumsdaine

CONTENTS

PREFACE iii

Chapter 1. The Analysis of Student Response as a Factor in Instruction. A. A. Lumsdaine 1

PART I. LEARNING COMPLEX SEQUENTIAL TASKS FROM DEMONSTRATION AND PRACTICE

Chapter 2. Theoretical Considerations in the Learning of Complex Sequential Tasks from Demonstration and Practice. Fred D. Sheffield 13

Chapter 3. Optimum Methods of Combining Practice with Filmed Demonstration in Teaching Complex Response Sequences: Serial Learning of a Mechanical-Assembly Task. Garry J. Margolius, and Fred D. Sheffield 33

Chapter 4. Combining Practice with Demonstration in Teaching Complex Sequences: Serial Learning of a Geometric-Construction Task. Walter Weiss, Nathan Maccoby, and Fred D. Sheffield 55

Chapter 5. Combining Practice with Demonstration in Teaching Complex Sequences: Summary and Interpretation. Nathan Maccoby and Fred D. Sheffield 77

Chapter 6. Repetitive versus Consecutive Demonstration and Practice in the Learning of a Serial Mechanical-Assembly Task. Garry J. Margolius, Fred D. Sheffield, and Nathan Maccoby 87

Chapter 7. Timing of Demonstration and Overt Practice as a Function of Task Organization. Garry J. Margolius, Fred D. Sheffield, and Nathan Maccoby 101

Chapter 8. Experiments on Perceptual Mediation in the Learning of Organizable Sequences. Fred D. Sheffield, Garry J. Margolius, and Arthur J. Hoehn 107

Chapter 9. Summary and Interpretation of Research on Organizational Principles in Constructing Filmed Demonstrations. Fred D. Sheffield and Nathan Maccoby 117

PART II. MOTOR, PERCEPTUAL, AND ACADEMIC SUBJECT-MATTER LEARNING

Chapter 10. Familiarization Procedures Used as Adjuncts to Assembly-Task Training with a Demonstration Film. J. Jepson Wulff, and Doris Kraeling 141

Chapter 11. Film-Mediated Learning with Varying Representation of the Task: Viewing Angle, Portrayal of Demonstration, Motion, and Student Participation. Sol M. Roshal 155

Chapter 12. Some Deleterious Effects on a Perceptual-Motor Skill Produced by an Instructional Film: Massing Effects, Interference, and Anxiety. William J. McGuire 177

Chapter 13. Some Factors Influencing the Effectiveness of Demonstrational Films: Repetition of Instructions, Slow Motion, Distribution of Showings, and Explanatory Narration. William J. McGuire 187

Chapter 14. Effects of Serial Position and Proximity to "Reward" within a Demonstrational Film. William J. McGuire 209

Chapter 15. "Response Guidance" as a Factor in the Value of Student Participation in Film Instruction. Gregory A. Kimble and J. Jepson Wulff 217

Chapter 16. The Effectiveness of Instruction in Reading a Scale as Influenced by the Relative Amounts of Demonstration and Problem-Solving Practice. Gregory A. Kimble and J. Jepson Wulff . . . 227

Chapter 17. The Effect of Animation Cues and Repetition of Examples on Learning from an Instructional Film. A. A. Lumsdaine, Richard L. Sulzer, and Felix F. Kopstein 241

Chapter 18. Factors Influencing the Effects of Student Participation on Verbal Learning from Films: Motivating versus Practice Effects, "Feedback," and Overt versus Covert Responding. Donald N. Michael, and Nathan Maccoby. 271

Chapter 19. Further Studies of Student Participation Procedures in Film Instruction: Review and Preview Covert Practice, and Motivational Interactions. Nathan Maccoby, Donald N. Michael, and Seymour Levine 295

PART III. STUDIES ON LEARNING OF PAIRED-ASSOCIATES MATERIAL

Chapter 20. Verbal Learning Efficiency as Influenced by the Manipulation of Representational Response Processes: Pictorial-Verbal and Temporal Contiguity Factors. Felix F. Kopstein and Sol M. Roshal 335

Chapter 21. From Audience Participation to Paired-Associate Learning. John Oliver Cook 351

Chapter 22. Response Analysis in Paired-Associate Learning Experiments. John Oliver Cook 367

CHAPTER 1

THE ANALYSIS OF STUDENT RESPONSE AS A FACTOR IN INSTRUCTION

A. A. Lumsdaine

Teachers as well as educational psychologists have long been familiar in a general way with the notion that students may learn more efficiently if conditions are arranged which permit them to rehearse or recite actively and appropriately the responses they need to learn. Student responses and the techniques by which they may be suitably controlled for effective learning are of both practical and theoretical interest. Their role in any systematic account of learning is of central importance.

The Role of Student Response

Disregarding for the present such questions as the exact functions of reinforcement and the precise mechanics by which unitary response acts are integrated into complex, sustained performances, it can be generally agreed that learning involves the development of new response capabilities that were not elicitable, prior to learning, by a given stimulus configuration or situation. This essential notion is applicable whether we are speaking of response "shaping" or are concerned only with bringing already well-defined responses under the control of new stimuli. Regardless of differences in specifics of interpretation, perhaps we may agree that, for the acquisition of new stimulus-response contingencies, a *sine qua non* is some form of occurrence of to-be-learned responses during the course of training—regardless of whether these responses are implicit or overt, simple or complex, unitary or integrated, symbolic or otherwise, mediating or terminal. In one sense the responses made by the learner thus comprise a crucial class of independent variables that influence the outcomes of learning experiences. In another sense, however, student response is never an independent variable, but only a mediating one. This is because we cannot manipulate responses *as such*; we can manipulate only the stimulus conditions that foster or regulate the occurrence of responses. In this sense, instructional control over the process of learning reduces simply and wholly to the control of stimulus conditions. But some selectivity must be exercised in choosing the stimulus variables on which attention is to be focused, and the writer believes that the most productive orientation for research which seeks to improve the effectiveness of instructional media is to identify those stimulus conditions through which appropriate responses of the student, *overt and implicit*, can be controlled effectively during the process of learning. To a considerable extent, the studies reported in this volume reflect this orientation.

Programmed Learning and Instructional Media

A few words are in order about the term "programmed instruction" used in the title of this book. It has both more general and more specific meanings. Generically, the notion of programmed instruction is perhaps nearly synonymous with that of "instructional media." This in general implies the presentation of instructional material to the student in a pre-planned, pre-sequenced order. In programmed instruction the sequence is determined or regulated either by the inherent structure of the media, by well defined procedures and constraints in the way they are used, or by the characteristics of mechanical arrangements which are used to present a "program" of material. Events instrumental in learning are thus generated which occur more or less predictably in terms of sequence and/or pacing. The sequencing and pacing need not be fixed, but variation with respect to them is constrained and is contingent on more or less clearly specifiable factors in the learning situation. Instructional media or programmed instruction provide, in other words, acts of instruction, not merely the wherewithal by which such acts might be effected by a teacher. The events generated in programmed instruction thus tend to be reproducible events.

It is this feature of reproducibility which makes instruction amenable to scientific study. With programmed instruction we can control and specify the experimental stimulus situation to a degree not often realizable when we seek to experiment with the properties of a "method" of instruction that is not embodied in tangible media. In the case of a film, for example, we can point to, display and describe very concretely what the experimental stimulus is. This is also true for an auto-instructional program designed for presentation in a teaching machine, or for a tape recording to be used in language instruction. The extent to which we can reproduce the stimulus material in constant fashion is, despite possibly quite important variations in the conditions and context of its use, still far greater than the reproducibility of the stimulus situation associated with the implementation of most so-called methods or procedures of instruction by a classroom teacher. It is obvious that such constancy, if not indeed a sine qua non for successful research on instruction, at least greatly facilitates such research.

Individual and Group Instruction

A more specific connotation for the term "programmed" instruction has evolved with the recent surge of interest (mainly subsequent to the conduct of most of the studies reported in this book) in teaching machines and auto-instructional (self-tutoring) programs. Many proponents of the latter would probably restrict the term "programmed instruction" to the special condition of learner-paced instruction that is properly regarded as one of the advantages of auto-instructional techniques. Without in any way deprecating the potentially great advantages of allowing each learner to proceed at a pace commensurate with his individual abilities, the self-pacing factor can usefully be distinguished from other aspects of programmed learning. As the writer has pointed out previously (Lumsdaine, 1959a), other major features of programmed learning are applicable both in the group and individual instruction situations. Outstanding among these other features, in addition to pre-sequenced, partially or fully automated presentation, are those of frequent active responding by the learner and of prompt feedback to inform him about the correctness or incorrectness of his responses. The latter features are those which figure prominently in the studies reported in this volume and which are implied by its title, "Student Response in Programmed Instruction."

Films (and televised instruction) commonly present instruction simultaneously to a large and/or spatially dispersed group of students. However, this is not necessarily the case; films and other sequenced visual (as well as audio) materials may, under suitable conditions, be presented to individual students. The recent emphasis on auto-instructional programs and devices (e.g., individual teaching machines and language laboratory installations) is closely bound up with the application of instructional media to individualized (or at least individual) learning. In a number of the studies reported in this volume (in Part I), filmed instruction was provided to individual students one at a time; in some of the studies reported in Parts II and III, however, group instructional situations were employed.

The fact of whether a particular study involved individual or group instruction may sometimes limit the applicability of the results to one instructional condition or the other, but it does not necessarily do so. Thus, studies that concern individually-varied versus fixed pacing are primarily applicable only to individual instruction, but studies involving cue or reinforcement factors built into a fixed instructional sequence have implications for instructional media generally, regardless of whether they were obtained in studies with individual subjects or in group instruction situations.

Experimental Study of Student Response as a Factor in Learning

Interest in experimental studies of the value of active student response, recitation, "participation" exercises, and the like derive in part from early studies by Witasek (1907) and Gates (1917). The advantage of active recitation over mere reading found by both forms a point of departure for a number of subsequent studies in which the role of active response by the student has been studied, and represents one of the historical antecedents for the work reported in this volume as well as for the current concern with self-instructional media that stress frequent active student response.

Gates compared the effectiveness of various allocations of total learning time between merely reading the material and attempted recitations, both for discrete, paired-associates material and for prose or "connected discourse" material. His data, as reproduced by Hovland (1951, p. 642), are shown in Table 1.

TABLE 1

Influence of Different Amounts of Recitation Upon Learning
(From Gates, 1917)

Percentage of Total Time Spent		Materials Learned	
In Reading	In Recitation	Syllables	Biographies
100	0	65.4	87.8
80	20	92.2	94.6
60	40	99.7	105.0
40	60	105.5	105.5
20	80	137.3	106.8

It is interesting to note that, although the amount learned increases progressively with increasing amount of time spent in recitation or active responding, the effect appears to be more marked for the isolated syllables than for the prose material (biographies), a finding also reflected in results of later investigators. In discussing these results, Hovland notes that Woodworth (1938) ascribes the relatively slight gains due to active responding in the case of prose material to the fact that in re-reading such material repeatedly the student is likely to anticipate what is coming and consequently to engage in implicit, or non-overt, recitation while reading. Such implicit recitation during the course of reading may of course occur during a first reading as well as later re-readings if the material is structured so that relevant gaps left in the material (as in the case of many auto-instructional programs) could readily be filled in, covertly or overtly, by the learner. The theory of implicit anticipatory responses is not only of considerable interest as a lead to the understanding of the nature of reading behavior, but also is important to the theory of learning from demonstrations, as developed by Sheffield in Chapter 2, and in accounting for other complex learning phenomena.

Student response is of general interest in programmed instruction for several different reasons. The first reason is that procedures fostering active, explicitly occasioned response (overt or covert) are generally, though not invariably, favored by experimental evidence over procedures which do not. A second reason is that when responses are made overtly they can be observed, checked on, and, potentially, more effectively controlled. The sufficiency of covert and/or implicit responding for mediating learning effectively is obvious, else one would be unlikely to gain much from reading a book without reading it aloud, from watching a demonstrational film without mimicing its actions overtly, or from listening to a lecture without chorusing back echoing responses aloud. But only by getting out a response overtly can we utilize its consequences directly in practice to guide the course of learning or to provide a constant check that it is "on course." It is the overt responses of the student which are utilized by a good tutor or coach in regulating instruction so that the right response is made, in guiding its elicitation until it appears, in providing specific, unit-by-unit reinforcement for successes, and in initiating prompt redirection of error tendencies.

At the same time, the effectiveness of implicit, non-overt responses in mediating learning can potentially be increased through better understanding of the way they function, and through stimulus-control measures designed expressly to take advantage of theoretically inferred implicit-response functions, even though these are not directly observed. The theoretical experimental work reported in Part I is of especial interest in this respect. It is certainly possible that stimulus-control features suggested by further analysis of the functioning of implicit responses may in the long run be of even greater practical importance in improving the effectiveness of instruction than the more obvious approach of manipulating conditions conducive to appropriate overt response. Finally, however, experimental conditions which permit study of the consequences of overt response may be of considerable indirect value in testing hypotheses about the functioning of responses which, in usual practice, remain at the covert or implicit level.

The Effectiveness of Overt-Response Procedures

The potential advantage of using overt-response procedures in employing instructional media is studied in a number of experiments reported in this volume. A prior, widely cited experiment was performed during World War II and was reported by Hovland, Lumsdaine, and Sheffield (1949). In this experiment, a film was used to teach groups of soldiers the military phonetic alphabet, Able for A, Baker for B, and so forth (plus some related information). Two forms of the film were compared in order to study the effect of using active review, or so-called "audience participation." The difference between the two forms was solely in the review sequences which followed the initial presentation of the letter-word equivalents. The control film used a standard "passive" form of review in which letters were presented along with their phonetic equivalents. In the active-review version, the letters only were presented and the audience members were instructed to try to call out the correct equivalents for each letter in turn. The fact that the correct equivalents had been recently presented provided the basis for their being emitted during the review. Feedback in the form of confirmation and/or correction (as well as prompting for the slower-responding trainees in the group) was provided by the fact that the whole group had to call out each response aloud. Since someone almost always gave a recognizably correct answer, prompts were almost always thereby provided for those who had not yet responded, and confirmation or correction was provided for those who had responded more promptly.

Data presented by Hovland, Lumsdaine, and Sheffield (also reproduced in Hovland, 1951, and in Lumsdaine, 1959a) showed consistent superiority of the active over the passive group at various criteria of recall-promptness in oral tests given at the end of the training. In preliminary work conducted in relation to a later replication of this experiment, Lumsdaine and Gladstone (1958) found that performance for the active group could be raised to near-perfect by increasing the amount of repetition of the response sequences. Further analysis of the data by Hovland, Lumsdaine, and Sheffield showed the interesting result that the difference in favor of the active-response and feedback procedure appeared to be least when least needed and most when most needed—that is, it was greatest for less-motivated, slower students in learning the more difficult portions of the material, and least for brighter, highly motivated students in learning the easier portions of the material. Some of the studies reported in the present volume (e.g., those by Michael, Maccoby and Levine) were designed to amplify these results by separating out the role of practice and motivation experimentally.

The experiment by Lumsdaine and Gladstone (1958) provided a replication of the earlier study by Hovland, Lumsdaine, and Sheffield, using the same subject matter, but differing in various aspects of experimental procedure and criterion measures employed. The results on active response verified the previous findings, indicating superiority for active-response procedure over passive review. In addition, the results showed that a simple, straightforward verbal presentation of the subject matter was more effective than one in which various audio-visual embellishments had been used in an attempt to make the presentation more interesting and to provide mnemonic associations.

In both of these experiments, the subject matter was verbal in nature and the task was a matter of simple one-to-one association between paired terms. Such tasks are of considerable prevalence and importance (e.g., learning a foreign language vocabulary—cf. the paper in this volume by Kopstein and Roshal—learning nomenclature, pairing names and faces, matching commodities and prices, etc.). Thus direct application of the method in many practical learning situations is possible, in addition to the more general theoretical interest of the results. The function of active-response procedures in insuring explicit rehearsal of correct associations is reflected in the considerable number of experimental studies performed since the conduct of the Lumsdaine and Gladstone study. Many of these are reported in the present volume. Both historically and, to some extent, conceptually, most of these studies grew out of the earlier studies by Gates and by Hovland, Lumsdaine, and Sheffield. In the main they were an attempt to follow up the implications of these studies, to investigate the variables governing the effects of student response in various instructional situations, and to progress toward better theoretical conceptualization of the role of both overt and implicit responding in these situations.

The applicability of the active-recitation principle is investigated experimentally in these studies for film-based teaching of various kinds of tasks, including the learning of facts and principles (e.g., the study by Michael and Maccoby, reported in Part II), of procedural skills such as equipment assembly (e.g., the studies by Maccoby, Sheffield, and collaborators reported in Part I), and of perceptual-symbolic skills such as reading of slide-rule scales (e.g., the study by Kimble and Wulff, Part II). From this experimental work has come considerable information on which to base principles concerning specific conditions that influence the effectiveness of active response interspersed in the context of expository or demonstrational instruction. The experiments reported in this volume thus extend our knowledge of factors important in the practical use of active-recitation or overt-practice sequences, but equally important is the light which some of them shed on the functioning of implicit responses and the stimulus conditions by which these responses can be made to operate more effectively in learning.

Some of the studies reported herein have recently been briefly summarized by Allen (1957, 1960), by Lumsdaine (1959a), and by Cook (1960a), though detailed accounts of many of them have not been previously available outside of government memoranda and reports with very limited circulation. The studies reported here derive in the main from a concerted program of research on student response factors, initiated and directed by the editor under Air Force sponsorship from 1953 to 1957. The general programmatic character of this work was initially described in a paper by Lumsdaine (1953), which also summarized briefly some of the earlier experimental results. The basic reports of most of the studies done in connection with this program have been included here. A few other related studies performed under this program, and a few studies deriving from other sources, are discussed in the concluding chapter of the volume and are abstracted in the appendix.

Organization of the Volume

Aside from the present introductory paper and the concluding chapter by the editor, the papers are grouped in the three main parts of the volume. A brief overview by the editor precedes each group of papers. Part I comprises eight chapters which report a series of studies centered on the learning of complex sequential tasks from filmed demonstration and practice. All of these studies were integrated around a central set of theoretical considerations described by Sheffield in Chapter 2. Part II presents a somewhat less integrated group of studies, concerned with the learning of various symbolic, perceptual, and manual skills, on which instruction was also given by films and related media. The studies in Part III are somewhat heterogeneous in origin and nature, but are all concerned with verbal learning, mostly in the form of paired-associates material. An attempt to draw together some of the common threads and implications from the three groups of studies is given in the concluding chapter, in which their relation to more recent studies of auto-instructional programs and teaching machines is also discussed.

PART I

LEARNING COMPLEX SEQUENTIAL TASKS FROM DEMONSTRATION AND PRACTICE

INTRODUCTION TO PART I

The studies reported in Part I represent a rather closely integrated effort carried out from 1954 to 1957 at Boston University. These studies, initiated and monitored by the editor and by A. J. Hoehn, were supported by two Air Force contracts as part of the program of the Training Aids Research Laboratory of the Air Force Personnel and Training Research Center. Nathan Maccoby, then Chairman of the Department of Psychology at Boston University (now Professor of Communications Research at Stanford University) was principal investigator for these contracts. Fred D. Sheffield of the Yale University Department of Psychology served throughout the project as major consultant, working in close collaboration with Dr. Maccoby in the design and analysis of the studies, and, in particular, formulating the main theoretical notions on which they were largely based. A particularly interesting and important aspect of this theoretical formulation is Dr. Sheffield's development of S-R contiguity theory to deal with the learning of implicit perceptual responses.

All of these studies dealt with the form of serial learning involved in learning to perform sequential or procedural tasks in which a lengthy series of steps must, by the nature of the task, be performed in a virtually fixed sequence. Such tasks, whether manual or verbal, are of importance in many fields of activity. Moreover, the implications of the research here reported are not wholly limited to tasks of this kind; some of the main theoretical considerations that apply to them, described by Sheffield in Chapter 2 and in later chapters, are evidently of considerable potential generality over a wider range of human learning, even though they are focused here on the learning of sequential tasks. For example, the studies reported in Chapters 3 and 4 and summarized in Chapter 5 have considerable relevance for the notions of small steps and "vanishing" of prompts as applied to the programming of academic subject matter (cf. Skinner, e.g., 1958), as has been previously noted by the editor (1959a). These and related implications are further commented on in the concluding chapter of the present volume.

Two main series of studies, corresponding to the work performed under two successive contracts, are reported in Part I. The first of these series is concerned, at the operational level, with the temporal distribution of alternated demonstration and practice and, more particularly, with variation in the length of segments of demonstration prior to attempted imitative practice by the student—a special case of the classical "whole-part" problem in learning. In terms of theory, as discussed by Sheffield in Chapter 2, this work bears on the competing requirements of temporal immediacy of demonstrational cues and attendant ease of practicing task segments correctly, versus the problem of integrating component responses into a lengthy sequential performance. As is made clear by Sheffield in Chapter 2, the relevant theory applies not only to the manual-construction tasks used in this experimental work but also to sequential verbal tasks such as the presentation of a speech and the calling out of relevant information on request on the basis of a "perceptual blueprint." The exposition of underlying theorization in Chapter 2 is followed by a detailed account of experiments, on two different tasks, presented in Chapter 3 (a mechanical-assembly task) and Chapter 4 (a geometric-construction task). The experimental data reported considerably amplify the summary report earlier given by Maccoby and Sheffield (1958). Chapter 5 presents a summary and integration of this series of experiments.

The second series of experiments, reported in Chapters 6 through 9, also reflects some of the theoretical notions presented in Chapter 2, but it is concerned with additional aspects of the temporal distribution of demonstration and practice segments, as well as with other experimental variables related to the underlying theory. These variables are closely related to concepts of perceptual mediation in the learning of "organizable" sequences. Two different mechanical-assembly tasks were used in these experiments. Chapter 9 provides a summary and interpretation of the experimental work reported in detail in Chapters 6, 7, and 8.

These two series of studies, both conducted by the same principal personnel and both drawing on common basic theoretical conceptions set forth in Chapter 2, represent a desirable trend toward concerted, programmatic effort in the study of factors governing the role of both implicit and overt student response in programmed learning. Unfortunately, shifts in mission emphasis and organizational changes of the sponsoring agency prevented the further continuation of this programmatic effort, which could have extended both its theoretical and practical implications. Of particular interest in further work which should follow up the promising leads provided by the work reported here would be the verification of experimental findings in other contexts and the conduct of experiments to test individually some of the principles, summarized in Chapter 9, which were used in constructing filmed demonstrations in this group of studies.

CHAPTER 2

THEORETICAL CONSIDERATIONS IN THE LEARNING OF COMPLEX SEQUENTIAL TASKS FROM DEMONSTRATION AND PRACTICE[1]

Fred D. Sheffield

A common training problem is the teaching of complex sequences of responses which must be learned in a fixed serial order. In such training situations, both demonstration and practice can be used. Demonstration is always necessary in order to show the learner what must be practiced. Practice is not always necessary along with demonstrations, although it is well established that some form of overt practice adds considerably to the effectiveness of training procedures. With simple or brief tasks, practice is not necessary because a single demonstration is sufficient to teach the perceptual and symbolic material necessary for immediate correct performance. In such cases some guided practice may help, but it is not an essential part of the training procedure, provided the learner is shortly given a chance to perform the task on his own while he still remembers the demonstration. Such subsequent performance actually constitutes practice even though it is not a part of the formal training procedure.

The need for formal overt practice as a supplement to demonstration becomes critical, however, when the task to be learned consists of a series of long and complex responses. What is learned from a demonstration is a sequence of perceptual and symbolic responses which must not only be learned in the correct order but must later be translated into correct overt performance. In the case of lengthy and complex material the attempt to practice correct performance may be defeated by the fact that the learner remembers so little of the demonstration that he does not know what to practice. The serial learning of the perceptual-symbolic sequence, which makes up his responses to the demonstration, may be so attenuated by conflicting memories that he is at a loss to utilize a subsequent practice period. Such conflict of recall, technically called "intraserial interference," is a common phenomenon in the learning of lengthy material.

---

[1]This chapter is primarily based on a report first submitted on November 27, 1954, under contract AF 18(600)-1210 between Boston University and the U. S. Air Force. It was subsequently reprinted for limited circulation as a Laboratory Note for the Training Aids Research Laboratory (Sheffield, 1956), and also as a Technical Memorandum for the Air Force Personnel and Training Research Center (Sheffield, 1957b). The latter portion derives from a second Technical Memorandum by Sheffield (1957a).

In such cases the learner would be unable to practice the correct sequence of responses after a single demonstration regardless of whether or not the overt practice were a part of the formal training procedure. If, on the other hand, the learner could take time out during the course of the demonstration to practice overtly whenever he reached a saturation point of what he could assimilate from the demonstration, practice might be utilized more profitably. In fact, if the task becomes sufficiently long and complex, some form of interspersed overt practice becomes a necessity if the learner is to profit much at all from the demonstration.

At the same time, however, stopping the demonstration in order to practice fractions of the task is likely to prevent the integration of the total task into the single sequence of responses that constitutes performance of the task as a whole. The method of interspersing practice can range all the way from practicing the entire task after each total demonstration down to practicing each small step of the task as it is shown in the course of the demonstration. In the former case the advantages of integrating the task are maximized but the advantages of clear recall of the demonstration material are minimized. In the latter case, on the other hand, the learner has clear recall of each little step, but is handicapped in integrating the steps of the total task. Thus each step of the demonstration is interrupted for practice, and each step of practice is interrupted for the next step of the demonstration.

This general problem defined the research area for the present chapter and the seven which follow. The present chapter will be concerned with the theoretical framework which served as the guide for the experiments and interpretations of results reported in Chapters 3 through 9. The organization of the chapter will be from general to specific, starting with the most basic assumptions and hypotheses and working up to the more concrete implications involved in the research to be reported.

Basic Theory

A fundamental assumption underlying all of the theoretical considerations is that learning *per se* requires only association by contiguity. This process will usually be referred to as "conditioning." A considerable portion of the theory is concerned with perceptual behavior and the mediation of overt performance by perceptual responses learned during "passive" exposure to demonstration materials. As far as learning is concerned, however, perceptual responses will be treated as following the same principles as motor responses.

Nature of Perceptual Responses

Sensory Responses. An important assumption is that of the existence of sensory responses which are completely central in locus and which need not have any motor components. Such sensory responses are assumed to be subject to the learning principle of association by contiguity and are assumed to have cue properties as well as response properties. That is, a sensory response can not only be connected to a cue, but also is a cue to which other responses can be connected. The assumption that such responses are entirely central may not be necessary, and perhaps an equally consistent theoretical development could be made on the assumption that all stimuli produce distinctive implicit motor responses, the cue property being the distinctive feedback from the distinctive implicit motor response. For the time being, however, it seems useful to assume entirely central sensory responses which are elicited independently of any motor response which might be made to the same stimulus—as if pain were sensed independently of hand withdrawal when one's finger is pricked by a sharp object.

Compatibility and Incompatibility of Sensory Responses. Sensory responses are expected to exhibit phenomena of compatibility and incompatibility comparable to those found in the case of motor responses, although exact analogies or identical principles would not necessarily be expected. Sensory "behavior" has not been well systematized in these terms, and there are few established principles. Also, the phenomena appear to differ from one modality to another—or even for different sensory qualities within a given modality. For example, in the case of color vision a perfect example of mutual inhibition can be found for complementary colors, as when yellow and blue light combine to produce no hue at all. At the same time, yellow and red combine to produce a compromise response—orange—which blends the two components in such a way as to produce a third hue which is similar to each of the two components. Such detailed ramifications of sensory interaction are not particularly germane to the present theoretical development. They are mentioned to bring out the fact that sensory responses involve questions of interference, facilitation, and response mixtures comparable to those studied at the motor-response level, including verbal responses.

Distinction Between Sensory and Perceptual Responses. As a rule, modern psychologists do not attempt to make a clear-cut distinction between sensation and perception. It will be convenient here, however, to make such a distinction in terms of whether the sensory responses involved are innate or conditioned. The word "perception" is generally used to refer to a process of interpretation of sense data, the word "interpretation" being used advisedly because it refers to an immediate "unconscious inference" and not a rational or verbal inference. It is also widely recognized that such "inferences" are highly dependent on past experience with the stimulus objects involved, especially at the human level. The position taken here is that what is usually called "perception" refers to cases in which the immediate sensory stimulation is not only eliciting its innate sensory responses, but is also eliciting other sensory responses which have been conditioned to the immediate stimulation in past experience.

Thus a block of ice presented only visually is perceived as cold because in the past it has been sensed cutaneously while being sensed visually. Similarly, a visually presented orange will be perceived in terms of its tactile, olfactory, gustatory, etc., aspects, if all of these have been directly elicited in conjunction with its visual aspects in the past experience of the individual. It should be clear that perceptual responses always involve sensory responses, the distinction being only that of whether some of the sensory responses are being "filled in" by means of the conditioning mechanism. The "filled-in" parts constitute the "interpretation" of the immediate sensory stimulation, as when one perceives an airplane from the distinctive sound made by its motor.

It is recognized that this restriction in the meaning of "perception" may not be very acceptable to those who argue against a distinction between sensation and perception or to those who carry out investigations concerned with whether or not a particular form of "perceptual" behavior is innate or acquired. The present position is that the distinction being made here is an essential one regardless of whether the more traditional sensation-perception question has an answer, and that alleged examples of innate "perceptual" behavior are better treated as innate sensory interaction, reserving the word "perception" for acquired connections between sensory responses.

Viewed in these terms, one of the most important features of perceptual responses is that they permit complete representation of a distinctive stimulus object, even though the object may be absent or even though all of the various stimulus aspects of the object may never be sensed simultaneously. Thus a physical object such as a whole orange may be immediately sensed only visually and at a considerable distance, but by means of perception its larger retinal image when held in the hand can be filled in, along with the bumpy and waxy feel of its surface, its odor, and its mass when "hefted" in the hand. Moreover, perception in effect makes the orange completely "transparent," permitting the sensing of the white insides of the skin, the many sections, the seeds, and so forth, even though the immediate sense data involve only a visually presented whole orange at a distance. In the same vein, a wristwatch is completely "transparent" to a skilled watch repairman. From the outside he can note the distinctive brand and model; this is sufficient for him to "fill in" all of the internal parts—their sizes, shapes, arrangements, and so forth. When he takes the watch apart, he is completely prepared for everything he sees because his anticipatory conditioned sensory responses correspond with his immediate unconditioned sensory responses when he opens it and makes the inner works visible.

Mechanics of Perceptual Redintegration

If a set of n sensory responses, R_{s1}, R_{s2}, R_{s3}, . . . R_{sn}, are elicited by different aspects of a given stimulus object, they will become conditioned to each other in the course of exploration of the object, which stimulates first with one aspect, then with another, and so forth, as it is examined, manipulated, sensed with different modalities, etc. Thus an object like an orange is smelled, touched, hefted, peeled, tasted, etc., giving rise to a succession of distinctive sensory responses which become conditioned to each other as cues. In the great variety of experience provided when a child becomes familiar with an orange, practically every stimulus aspect has sometimes preceded, sometimes followed, and sometimes occurred simultaneously with every other aspect, giving rise to a conditioned (perceptual) response pattern which is unique for oranges as objects and which can be elicited in relatively complete form by only one unique aspect of the orange. If we simplify the real orange to a four-aspect object this can be diagrammed as shown in Figure 1.

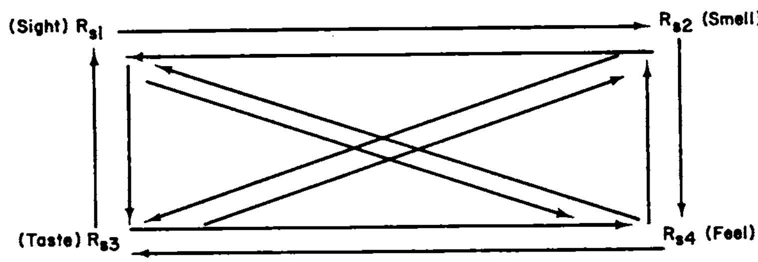


Figure 1. Organization of Sensory Responses to Different Aspects of a Stimulus Object into a Single Perceptual Response.

This "cross-conditioning" mechanism accounts for the "filling-in" property of perceptual behavior in which a fragment of a total stimulus-pattern "redintegrates" the whole. (The term "redintegration" is borrowed from Hollingworth (1928), who used it in much the same sense.) This mechanism also accounts for the "all-or-none" property of most perceptual responses: since all of the sensory responses—as cues—are conditioned to each other, the perceptual response tends to be elicited as a whole once it is initiated, and further tends to maintain itself once it has been elicited as a whole. That is, each sensory response serves as a cue which feeds into the others, making the total response stable and temporarily resistant to any interruption from new stimuli.

It should be noted that the total perceptual response at any given moment will not necessarily include all of the possible cross-conditioned elements. Some of these may be incompatible with each other, so that while one conditioned sensory response is being elicited, others might be impossible of elicitation. Also some conditioned sensory responses might be incompatible with innate sensory responses being elicited at the moment. The expectation would be, therefore, that the "total perceptual response" would be fairly labile, shifting from moment to moment in the particular constellation of conditioned sensory responses which constitute the momentary perception of the object. With a complex object, however, it would nearly always be expected that a sufficient number of different sensory responses would be elicited at any time to make the momentary perceptual response unique for the object. Thus at any particular moment the expert watch repairman's perception of the inner works of a watch might include or exclude the escapement device. In general, however, the particular set of sensory responses he did have at any moment would be sufficient to provide a constellation unique to the actual object, distinguishing, for example, watches from clocks, and one make of watch from others.

Another feature of perceptual responses is that a complete perceptual response can be elicited by a conditioned stimulus in the absence of any of the stimulus aspects of the perceived object. Perhaps the most common case is that in which the object is given a verbal label, such as "wristwatch." This label may be seen in writing or heard as a spoken word. In neither case does it provide any of the stimulus aspects of the object except through conditioning of the perceptual response of the object to the completely unrelated stimulus aspects of the label. Thus an entirely neutral cue may become a source of a complete perceptual response to a totally absent stimulus object. The point is that not only do perceptual responses "fill in" the missing parts of the object when it is presented only in fragmentary form (as when one looks at a watch but cannot see the works), but also perceptual responses appropriate to an object can be present even when the object is not present.

Still another feature of perceptual redintegrations which is implied by the foregoing is that they can serve as stimulus patterns as well as being elicited as organized response patterns. Since the sensory-response components of the perceptual response have stimulus properties, and the pattern of the response is a unique and relatively stable one, a well-established perceptual pattern (such as the perception of an orange) can serve as a highly distinctive cue-combination which can be conditioned as a unitary stimulus to any other response. Thus a child learns to say "orange" when he perceives an orange, and naming based on the visual orange should transfer to the olfactory orange, the gustatory orange, etc., if the child had sufficient prior familiarization with oranges as real objects.

Integration (Organization) of Response Sequences

What has been said thus far has been oriented toward the formation of perceptual "gestalts" through experience, the chief assumptions being association by contiguity and the response and stimulus properties of sensory events. It should be noted that in the present use of concepts like "organization," "structure," and "gestalt" in speaking of the end products, it is to be understood that a theory of what these are and how they are formed is being considered; structure is deduced, not assumed.

The same sort of analysis as has been described would also apply to overt motor responses in some instances, as in posturing, grimacing, holding a note while singing, etc., where the emphasis is on a stationary pattern of complex muscular contractions which holds itself together as a unit by the "cross-conditioning" of feedback cues from the various muscles. The analysis is more typical of sensory responses, however, if only for the reason that a sensory response is a type of response which begins with the onset of the stimulus, for the most part stays the same as the stimulus persists, and ends when the stimulus (which may persist a long time) is terminated. This is not characteristic of overt responses, which commonly involve successive movements, the initiating of cues for each successive response being removed by the execution of the particular response.

A different kind of analysis of response organization applies when responses are necessarily sequential. This analysis is outlined below and applies to sensory and perceptual responses as well as to motor responses.

If the sequence of responses, a, b, is elicited by two independent stimuli a number of times, there will be a tendency for the cues arising during the performance of a to become conditioned to b, such that with appropriate motivation the conditioned sequence, a—b, will tend to be set off by initiating a even without presenting the originally adequate stimulus for b. This is the smallest element in a conditioned response chain.

Furthermore, if b', a part of response b, can occur simultaneously with a, it will tend to be elicited by cues during performance of a and prior to complete performance of b. This b' portion of b thus moves forward to the beginning of the chain, and the resultant pattern can be diagrammed as indicated in Figure 2.

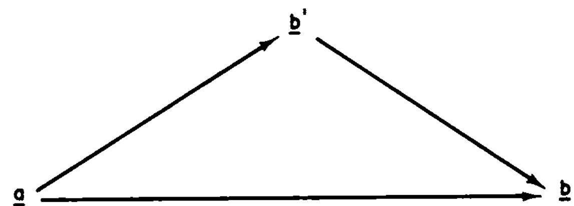


Figure 2. Integration of a Sequence of Two Responses.

This smallest link in the chain is now integrated by the portion of b that moves forward. Part of the subsequent response is present in the performance of the antecedent response, and cues from both a and b' are conditioned to performance of b. The response is also coordinated in that, instead of two discrete responses occurring one after the other, a blending of the two responses occurs in which all of b that is compatible with a occurs along with a, and a graded transition from a to b characterizes the execution of the link. This derivation assumes only the commonly reported tendency for conditioned responses to move forward in time if CS precedes US by an interval longer than the latency of CR [cf. Hilgard and Marquis, 1940, 1961].

If instead of the two responses, a, b, there are three responses, a, b, c, independently elicited in a regular sequence, the same relations as outlined above will hold between a, b, and b', and between b, c, and c'. In addition, however, if c'', a portion of c', can occur simultaneously with a and b', it will move forward by higher order conditioning to the beginning of the sequence, giving rise to the pattern indicated in Figure 3. This is also an integrated and coordinated sequence as in Figure 2, and merely involves a more complex pattern of cue control in the unification of the discrete original responses into a smooth-flowing series. In general, c'' would be a weaker version of c than c', because c' would have to be compatible with only a single response, as in the compatibility of c' and b.

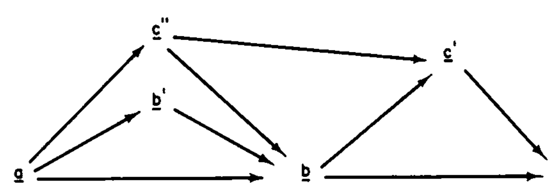


Figure 3. Integration ot a Sequence of Three Responses.

Descriptively, then, the integration of a sequence is a process of a moving forward in the sequence of all of those later parts which can occur simultaneously with earlier portions. This process results in a sequence more stable than a mere chain such as a→b→c. The reason it is more stable is that when a is initiated, b and c are immediately aroused in incipient form, which gives them prepotency over any incompatible responses which might be stimulated by situational cues arising during the performance of the sequence.

A more complex but analogous integration is expected in any still longer sequence: a, b, c, d, . . . n. A good example of such a sequence is the pronouncing of a familiar word. The whole word forms "on the tip of the tongue"; anticipatory portions of subsequent vowels and consonants are present as pronouncing begins, and early portions shade into later portions by imperceptible degrees as the word is spoken.

The properties of such integrated sequences implied thus far are:

1. They behave as a unit which is tripped off as a whole.

2. The elements of the sequence are resistant to associative inferences once the unit is initiated. We may also infer a third proposition, namely, that:

3. Integrated sequences are less stable the larger the number of elements, n. This third property follows from the already noted inference that the longer the sequence the smaller the portion of the terminal response that can move forward, since such remote anticipatory responses must be compatible with an increasing variety of responses the further forward they move. Other implications concerning integrated sequences may also be inferred. One important one is that:

4. Sequences are more stable and more quickly integrated the more compatible the separate elements of the sequence. This follows from reasoning similar to that in proposition 3 above. Further, we may deduce that:

5. Sequences will be more quickly integrated with repetition if changes in context cues from a to b, to c, etc., are relatively slight during the sequence but markedly different before and after the sequence.

That is, if a common set of cues other than those from sequential responding tends to be present during the execution of the sequence, this common context will serve an integrating function comparable to that of c" in Figure 3. Such a constant context would tend to elicit incipient performance of all responses in the sequence. The constant context would be particularly effective if a lengthy sequence were divisible into sub-sequences, each having a different context. This would aid integration within a sub-sequence and help prevent interference between elements of different sub-sequences. As an illustration, the outline headings in a lecturer's notes for a partially memorized speech serve to integrate the topics to be recited under each heading.

It should be understood that this analysis of serial integration applies to any kind of regularly practiced response sequence. It covers the smoothly executed responses of a well-coordinated motor habit but also applies to perceptual sequences where all of the responses are presumably central.

Superordinate Sequences

One implication is that a long sequence of separate elements cannot in itself become integrated. The limit is set by how far forward "incipient" response elements can move, and whereas the execution of a given part of the sequence involves integration of a segment of about-to-be-performed responses, this control does not extend clear to the end of the sequence. A current distracting stimulus may therefore initiate responses incompatible with some of the later elements, and they will not be prepotent when the time comes for their performance. Such a situation rarely exists for long, however, because the sequence tends to break down into natural units which themselves become integrated with practice. These natural units become "elements" in a new sequence and reduce the number of such "elements" in the total sequence to a smaller number. These can also become integrated in the manner alluded to above. The speaker with an outline of six headings can memorize the headings as an integrated sequence not subject to interference, and if the material under each heading is further subdivided into six subheadings which are themselves individually integrated, a compact structure of his talk is created and it goes off in correct sequence despite distractions. It does not go off entirely as a chain, and it does not go off as a single integrated sequence. Rather, it is elicited as an integrated superordinate sequence which cues off integrated subordinate sequences. We can diagram this in the simplest case as follows, letting arabic numerals stand for a superordinate sequence.

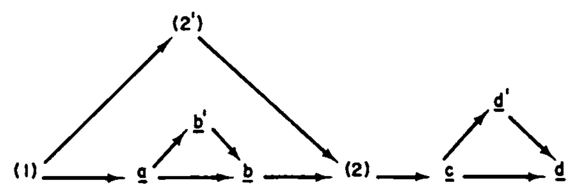


Figure 4. The Mechanism of the Integration of Lengthy Sequences Through the Formation of Superordinate and Subordinate Sequences.

In practice a four-element sequence, a, b, c, d, would ordinarily be short enough to integrate itself with a structure like that in Figure 3. Figure 4 omits this structure and illustrates the mechanism involved in the formation of superordinate sequences as a structure which integrates a more complex case of a lengthy sequence of smallest elements. The mechanism probably often operates even in the simplest sequences, however, and in line with the four-element example in Figure 4 an absent-minded professor might find it useful to organize a four-item grocery list into two items of "meat" and two items of "produce" as a means of carrying the whole list in his head at once without interference from other things on his mind.

In the above discussion the concept of a "natural unit" was used. Empirically these are units that readily form a stable, integrated segment of the total sequence. Theoretically they are units for which incipient later responses in the segment can readily move forward to earlier portions, which implies the presence of either or both (a) a high degree of compatibility of responses, and (b) integrating context cues that remain fairly constant during the segment but which differ from the contexts of preceding and succeeding segments.

Relation Between Perceptual and Symbolic Responses

In the present theory a "symbol" is an arbitrarily determined stimulus which has been conditioned to the perceptual response appropriate to a stimulus object. A memory image of a person would not be a symbol, but the name of the person is a symbol, whether it is heard or seen in printed form. The reason for saying "arbitrarily determined" is that the sine qua non of the functioning of symbols in thought and communication process is that they are a code system which can be used in the absence of the real objects referred to. Like other code systems, they commonly stand for the thing represented by virtue of conventional rules requiring translation into and out of the code system. The name of a person is an arbitrarily determined property of the person, but, by virtue of practice at hearing, seeing, speaking, etc., the name in connection with the person, it becomes capable of redintegrating the perceptual response learned to the person as a stimulus object.

This treatment is not particularly conventional, and some writers would prefer to think of perceptual representations as "symbols" of the external thing represented. A distinction needs to be made, however, between perceptual representations, which are isomorphic with the object's sensory effects, and symbolic representations, which are arbitrarily determined equivalents established by society for purposes of communication. Thus there is absolutely no inherent similarity between the printed or spoken word "orange" and the physical object which it represents. It is only through the deliberate association of the word and the perceptual response to the object that the word takes on the capacity to elicit the perceptual response. Such associations are produced by formal training and should be distinguished from the perceptual learning which is forced by the physical environment when a child learns, for example, that an orange contains seeds and juice underneath its skin.

The most important symbol system in human behavior is, of course, language, and the argument concerning the arbitrariness of symbols might be said to break down in the case of words with onomatopoeia or in the case of the hieroglyphic picture-writing of some ancient cultures. Even in these cases, however, it is conventional usage which established the meanings and it is only in the process of evolving language that some words (or scribings) have some isomorphism with the object for which they stand.

A complete discussion of symbolic behavior is not relevant here, but a critical feature of such behaviors of concern in the present studies is that they can be manipulated by the individual in the form of symbolic responses (again chiefly language). By the use of his speech apparatus he can describe objects and events in words, memorize his verbal description, repeat it back to himself or others later, and so forth. He has a verbal code system which he can manipulate to recreate perceptual responses in himself—and in others as well.

Perceptual Mediation of Overt Sequences

Perceptual Serial Learning and Mediation. The type of learning task to be considered in succeeding chapters (3 through 9) is that of serial learning. Most emphasis will be placed on the serial learning of mechanical assemblies, that is, sequences in which mechanical parts are fitted together in the proper order to produce the completed mechanism. References will be made, however, to other types of serial learning, where relevant, to the general problems of organization and perceptual mediation. The major thought underlying the notion of "perceptual mediation" in learning is that learning can proceed at a perceptual level without any actual performance of overt responses. In this notion it has been assumed that perceptions are responses which can be connected to cues by much the same associative processes which apply to overt responses. It is also assumed that perceptual responses have cue properties, such that other responses—perceptual or overt—can become connected to them as stimuli.

This general approach allows for the serial learning of perceptual sequences independently of any overt responding. Thus, if a series of N perceptual events is repeatedly experienced, a perceptual sequence,

$$P_1 \rightarrow P_2 \rightarrow P_3 \rightarrow \cdots \rightarrow P_{N-1} \rightarrow P_N,$$

will be established. This type of perceptual sequence is of obvious importance when a person is learning from a demonstration, in which he does no overt manipulation himself. Assuming, however, that all of the manipulations he observes are those that he can already perform upon seeing them (e.g., use a screw driver), he will be able to reproduce the overt sequence represented by the perceptual sequence learned from the demonstration. This can be symbolized as follows:

$$\begin{array}{lcccccccccc} \text{Perceptual Sequence} & P_1 & \rightarrow & P_2 & \rightarrow & P_3 & \rightarrow \cdots \rightarrow & P_{N-1} & \rightarrow & P_N \\ & \downarrow & & & & \downarrow & & \downarrow & & \downarrow \\ \text{Overt Sequence} & R_1 & & R_2 & & R_3 & \cdots & R_{N-1} & & R_N \end{array}$$

In this representation the learner has not acquired any serial chain connecting the overt responses; rather, the overt responses are being elicited by the acquired perceptual chain. Thus the P's in the above diagrams stand for perceptual responses with cue properties. They therefore can both be connected in a serial chain, and be the cues for a sequence of overt responses. The overt responses have not at all been connected as a chain by the learner's experience, but they are elicited in the proper sequence because (a) the P's are connected in the proper sequence, and (b) the learner is already able to translate the perceptual responses into the appropriate overt behavior. This type of analysis applies in cases in which the learner observes a demonstration of the correct overt sequence and is subsequently able to reproduce the overt sequence from his memory of the demonstration.

Perceptual "Blueprints". This general approach also allows for another type of sequentially correct overt performance—again, without the formation of an overt serial chain. This second situation requires perceptual mediation but does not require the acquisition of a perceptual sequence. An example might be the situation in which a child is shown three colored blocks arranged in the order: red, white, and blue. If the child is shown the three blocks and asked to reproduce the arrangement, it need not have learned the perceptual sequence,

$$P_{Red} \rightarrow P_{White} \rightarrow P_{Blue} .$$

Rather, it may merely learn the pattern such that it can reproduce the visual image as a whole, in which the red block is on the left, the blue block is on the right, and the white block is in the middle. If the child can reproduce this whole pattern perceptually, it can again read off the correct overt performance, although in this case there is no need for chaining of a perceptual sequence. Instead the child matches its memory for the perceptual pattern against the perceptual pattern produced by its overt manipulation of the blocks, the overt efforts coming to an end when the two perceptual patterns coincide. This result may be symbolized as follows:

Memory Pattern:	Left P_1	Middle P_2	Right P_3
Immediate Pattern:	Left P_1	Middle P_2	Right P_3

where it is understood that considerable overt trial and error may have been necessary to get the final match.

This latter mechanism of perceptual learning may be likened to the situation in which a person constructs a mechanism from a blueprint. A blueprint has the advantage over perceptual memory of being exact, unchanging, and not subject to possible faulty memory. The individual can refer back to the blueprint innumerable times while he attempts to match his perception of his overt product with his perception of the blueprint. The mechanics are the same, however, whether the person matches his behavior to an external blueprint or whether he matches it to a memory image. The memory image is like a "blueprint," and the learner continues to manipulate until his perceptions of his immediate product match his perceptual memory.

Such perceptual units as the above play an important role in organizing some response sequences. In learning a mechanical assembly task, for example, the perceptual responses to the finished product may be organized into such a unit, which, when redintegrated by parts of the completed stimulus object, serves as a perceptual "blueprint" for reassembly. The learner can see at all times what he has assembled thus far and can "read off" from his perceptual "blueprint" what should be done next, even if he has not yet mastered the response sequence. This sort of an organizing or integrating factor may be of particular significance in learning from demonstrations, where overt performance is generally not practiced during acquisition, and where what is practiced or acquired must be subsequently translated into overt responses. At the animal level, Tolman (1948) has proposed the concept of "cognitive maps" as a kind of perceptual guide used by rats in mastering mazes. This conception is almost the same as the "blueprint" conception formulated here except that the use of the word "cognitive" suggests a process not implied by perceptual redintegration per se.

Related Hypotheses

Most of the foregoing theoretical section has been devoted to specifying the basic mechanics of perceptual behavior and its general role in acquisition of overt sequences. In the present section some related hypotheses considered more dubious will be outlined.

Telescoping of Time in Perceptual-Response Sequences

One such hypothesis is that a perceptual sequence which represents an overt sequence can be executed much more rapidly than the overt sequence. This hypothesis seems extremely probable if only because overt sequences require external manipulations and have physiological limits in the maximum possible speed of motor execution. Perceptual sequences would also have physiological limits, but being primarily neural, they should be capable of more rapid execution. Moreover, a perceptual sequence might omit many steps of the overt sequence without doing violence to the representation. For example, the actual ascent on foot of a high mountain might consume hours, but the perceptual representation of the ascent may take only a fraction of a second, or a few seconds if some of the steps of the ascent are taken into account.

One implication of this hypothesis is that a perceptual sequence which represents a lengthy overt sequence might be executed in a shorter time and with fewer actual steps. Even a complete sequence would go faster because of its neurological nature, but such a sequence could also collapse some of the overt steps as indicated above. If prior learning were sufficient for the learner to translate such "collapsed" perceptual sequences into appropriate overt behavior, the total task would be simplified in the number of units to be serially learned. For example, a neophyte in a machine shop might have considerable difficulty with the task of drilling a hole of a given size if he has had no prior experience. He must learn how to select the drill size, how to make a center-punch in the right place, how to turn on the drill, how to manipulate the drill press so that the hole is made in the right place, and so forth. This entire operation is greatly foreshortened, however, for the experienced shop man. He only needs a perceptual "glimpse" of the requirement that at this stage of the total task a hole should be drilled. His perceptual response may be in terms of the finished hole, the shape of the drill press, the appearance of the rack on which drills are kept, or any one of a number of different possibilities. But for him drilling a hole is not in itself a serial learning task. He has already mastered the sequential aspects of the task and can "read" from his perceptual response, which involves drilling a hole of a given size in a given place.

"Reading" an Overt Sequence From a Stable Perceptual Complex

Another hypothesis is that a correct overt sequence can be cued off by a unitary perceptual response which represents an entire sequence in completed form.

An extreme case might be that of the rare college student who still possesses "eidetic" imagery. Such a student can review briefly a mathematical proof shown on a page of a mathematics textbook, and then reproduce exactly the steps of the proof. Such an individual merely reads off the symbolic steps from his visual image. Apparently, if he has had prior experience with writing the symbols, he has little trouble in reproducing the steps even if he does not comprehend them.

The actual existence of such an extreme example may be doubted, but the example serves to illustrate the mechanism being described. The most extreme form is that in which the learner has a perfect memory image of the finished product and needs only to match his perception of his own product with his memory image. Most adults do not have eidetic imagery, but to a lesser degree most adults are able to match their overt behavior—and particularly the perceptual outcomes of their overt behavior—with their perceptual memories. As mentioned earlier, the mechanics of this use of perceptual memory are comparable to those of the craftsman who follows a blueprint or other diagram of a structure that is to be produced. The production of the structure involves a sequence of steps in terms of overt action, but the blueprint or diagram is a static complex which may be constantly referred to in executing the steps represented. In similar fashion a perceptual representation of a finished product may serve as a "blueprint" which guides the steps of the overt sequence.

Compatibility of Perceptual Versus Overt Responses

Another hypothesis is that as a general rule there is less interference between perceptual responses than between overt responses. From the present formulation, in which sensory and perceptual responses are given most of the same properties as overt responses, there would be no particular reason to expect that perceptual responses would be less subject to interference than motor responses. It might be expected that different sensory modalities would not interfere and that different retinal areas would not interfere. It may further be the case, however, that perceptual responses inherently interfere less with each other than do motor responses. To take a familiar example, most children get exposed to the trick of making a patting motion on the head with one hand while executing a circular motion on the abdominal region with the other. It takes only a few seconds to demonstrate the trick and leave the child with a perfect perception of what is to be achieved, but it may take many minutes of practice before the child is able to reproduce in his own motor behavior what he has seen.

Thus there may be a fundamental disparity between the interference factors at the overt level and the interference factors at the perceptual level. Speaking figuratively, the right hand may be obliged to know what the left hand is doing, but the "actions" of the right side of the retina may be fairly independent of what the left side is "doing." Also, it is expected that perceptual responses will tend to be non-interfering with motor responses which do not involve incompatible sensory responses.

Regardless of the correctness of the hypothesis that sensory (and perceptual) responses are less subject to interference than motor responses, the fact remains that complex perceptual responses can exist as static patterns which remain fairly stable while the individual attends to the different parts and guides his overt behavior accordingly. Thus a child may name all of the states with the help of a visual image of the map of the U.S.A. or a university president may name all of his professors by reference to a visual image of the campus and the office spaces in each of the buildings. Similarly, an infantry division commander might enumerate all of his officers above the rank of captain by reference to a perceptual image of his table of organization, which requires him to name an officer for each separate regiment in the division and for each separate battalion within each regiment. At the anecdotal level, at least, it appears to be the rule that such perceptual "images" can be reinstated as fairly complete units and without interference from the separate parts.

In the example of the division commander, his recall of his officers will not ordinarily be in a fixed sequence of verbal response. He might "read" his perceptual image of the table of organization either from bottom to top or from top to bottom. Or he might "read" from left to right or right to left. Such considerations would determine whether he gives colonels first and majors second, or vice versa, and whether he gives officers of low-numbered regiments and battalions before high-numbered ones, or vice versa. The main point is that he has a fixed image of the table of organization and can not only reproduce the name of each officer who occupies a place in this image but can reproduce these in any order requested.

Symbolic Sequences

An overt sequence may have its counterpart at the symbolic level as well as at the perceptual level. Going back to the earlier example, a child may reconstruct a red, white, and blue pattern of blocks on the basis of his visual image of the color pattern, but if he has learned words for colors, he may remember the verbal sequence, "red, white, and blue," and use this symbolic representation as the basis for his correct reproduction of the overt responses. His symbolic representation in such a case has the weakness that it is unclear as to whether to "read" from left to right or from right to left—a weakness that would not be present in his visual image.

The hypothesis assumed here is that symbolic behavior is relevant for its ability to reinstate a perceptual response. This emphasizes the communication aspect of symbolic behavior and ignores what may be equally significant aspects from other standpoints. For example, in the field of mathematics the most important feature of symbols is that they can be manipulated according to axiomatic rules. In such manipulations the symbols may have perceptual significance only at the beginning and end of the manipulations, whereas the complex manipulations themselves may be perceptually meaningless despite the fact that they obey the perceptually axiomatic rules of manipulation. The hypothesis also raises the question, however, of whether a fixed sequence of symbolic responses can be more readily learned than a fixed sequence of perceptual responses. If this were the case, then a perceptual sequence could be most readily learned by learning a symbolic sequence such as a series of words, each of which was cued in past experience to the appropriate perceptual response. The appropriate perceptual responses would fill in the details needed to translate the sequence into overt action, and the readily learned symbolic sequence would be responsible for the production of the appropriate perceptual responses in the appropriate order.

Investigation of this problem has not been reported in the literature, and no definite stand on the issue can be taken here. It is worth noting, however, that in paired-associates learning the connection between stimulus word and response word is usually more readily established if the learner gets what he calls an "association." By this he usually means that he has found a perceptual image which includes both symbolic responses—as when he "associates" the words "hot" and "far" by thinking of the sun, which is a single object having both attributes. The present assumption is that perceptual responses are learned at least as easily as their symbolic equivalents, but it should be understood that this is an uninvestigated research area.

Role of Perceptions in Organized Sequences

Any serial-learning task is organizable into sub-parts which make up the whole. The organization may be entirely arbitrary and externally imposed, as when a long list of nonsense syllables is subdivided into two halves, four fourths, etc. Or, the organization may be inherent in the sequence, as when a mechanical assembly is subdivisible into sub-assemblies. It is here assumed that all forms of organization provide a mnemonic advantage in serial learning. The distinction between imposed organizations and inherent organizations is probably an important one, although organization in either case is assumed to aid memory.

A simple example of an "imposed" organization is seen in the common method of teaching the alphabet to youngsters by having them sing the letters in a song which incorporates the words into an auditory pattern having meter, rhyme, and melody. Presumably the melody is easier to learn than the alphabet, and since each letter has its own distinctive cue in the melody, it is easier to recall that letter at the right point in the sequence. Additional advantages of the song are that it has meter, which helps insure that none of the letters is left out, and rhyme, which helps elicit the right letters at the rhyming points, which in the alphabet song are at G, P, V, and Z. The advantages of melody, meter, and rhyming points are important for the special case of memorizing song lyrics but otherwise have no general significance in the psychology of learning. This follows a general rule that a major mnemonic advantage of organization in serial learning is the distinctive context-cue provided by the separate parts of the organization.

A simple example of "inherent" organization is found in the case in which the medical student masters the names of all the bones in the human body. He can break the human skeleton down into "axial" and "appendicular," he can break "axial" down into "cranial," "facial," "trunk," etc., and he can break "appendicular" down into "upper extremities" and "lower extremities." Or he may find it useful to break "trunk" down into "thoracic" and "pelvic." In any case the parts are inherently organized, and it would generally be expected that he can recite them better by utilizing headings and subheadings than he could by attempting to memorize them as an undifferentiated sequence.

The role of perceptual behavior in the mnemonic advantage of organized sequences is that of providing a distinctive stimulus context within which appropriate response items can be specifically cued. Thus the medical student will be less likely to think of "femur" and more likely to think of "humerus" if the perceptual context is "upper extremities"—despite the similarity of these two bones on other bases (i.e., they are single bones in the "upper" part of the two kinds of limbs). The same analysis cannot, of course, be made in the case of the alphabet song, since this is a case in which the context is external rather than perceptual. That is, the child actually hears his melody, meter, and rhyme, whereas the medical student must perceive the human body without having one to look at. It is obvious, however, that the stimulus-response mechanics are identical in either case.

The importance of the distinction between imposed and inherent organizations is that memorization is simplified in either case, but, whereas mnemonic simplification is guaranteed if the organization is inherent, it may or may not be obtained if organization is not inherent. In other words, a memory task may be simplified by imposing an organization on something not inherently very well organized. Thus, if one's purpose is to facilitate memorization, one should attend to the various ways in which organization can be imposed on an otherwise poorly organized set of materials.

In the perfect case of a maximally desirable training procedure, the imposed organization would be entirely perceptual, as opposed to some overt ritual (as in the case of the alphabet song for children). This would ensure that no external "props" were required to elicit the correct sequence. Also, the most desirable procedure would capitalize on any organizational factors inherent in the task itself. Beyond this the optimum procedure would impose an additional perceptual organization which would permit the learner to memorize the sequence by means of an easily learned perceptual superstructure which provided distinctive cues for each element in the overt sequence.

"Natural Units" of a Sequential Task

The concept of "natural units" of a sequential task has been mentioned several times earlier, but since this concept plays a fair-sized role in the research to be described it deserves further discussion now that the basic theoretical considerations have been covered.

A "natural unit" may be defined as a segment of several or many elements of a sequential task which, because of their inherent properties, are fairly readily integrated into a superordinate unit of the total task. They are parts of the "inherent organization" alluded to in the previous section. Memorizing the alphabet is a learning task with no natural units, even if it can be "organized" by imposing an extraneous structure which divides it into the aforementioned metrical and melodic units ending at each of the rhyming points. Natural units have, by contrast, beginnings and endings dictated by the task and can usually be recognized a priori by the following properties:

1. A common and distinctive context, which is quite different from contexts of other parts of the task. When these conditions are met, the steps within a context are cued to the context and protected from interference from steps practiced in other contexts.

2. A context which suggests the steps as well as being distinctive. If a mechanical assembly, for example, involves a gear box transmission in a single housing, the context is not only uniform but sets the learner to look for gears in seeking each next response in the total assembly, until the unit is completed and the context changes.

3. Adaptability to the superordinate sequence which permits collapsing the parts of the unit into a smaller number of sequential items that do not interfere. From this standpoint a natural unit begins and ends when prior and subsequent elements in the total sequence cannot be easily incorporated into a superordinate sequence.

4. Susceptibility to perceptual "blueprinting" of the type mentioned before, in which the segment can be perceived "all at once" and the elements of the segment read off in sequence.

From this it should be clear that the whole task is itself a natural unit, and that if it is at all organizable it can be broken down into smaller and smaller natural units, like the outline of a well-organized speech. The "natural units" in the studies to be discussed are of the larger variety, that is, they are the largest sub-units into which the total task could be subdivided.

Research Program

As will be understood, the nature of the research reported in the following chapters was not of the form of direct tests of the basic theoretical considerations expounded above. The theory guided the research but the research was not designed simply to test the theory. As in most military research, certain practical issues were at stake and the role of the theory was mainly to direct the investigations into psychologically meaningful and generalizable channels.

As noted at the outset of this chapter, the basic practical issue was that of how to combine demonstration and practice in the case of lengthy sequential tasks, in which it was fairly obvious that a single demonstration of the entire correct sequence would leave the learner in a state of memorial confusion in which he would be unable to practice the correct responses even when given the opportunity. If he cannot even remember the perceptual and symbolic sequence (which he passively received) he will be poorly prepared to practice overtly. The implication was that "time out" for practice should be taken at points when effective practice was still possible, i.e., when a demonstration segment was still well enough remembered that it could be translated into overt practice.

This general problem gave rise to the concept of the "demonstration-assimilation" span (D-A span), comparable to the "memory span"—or "span of apprehension"—of classical studies in perception. For a particular learner a D-A span would ideally be the amount of demonstration material he could be exposed to before his ability to reproduce it correctly fell below 100 per cent. Since different individuals differ in this respect, a compromise definition of a "D-A span" was used, namely, the length of a segment which could be reproduced perfectly by 75 per cent of the subjects in preliminary testing.

The D-A span breakdown of any lengthy sequential task (taught by demonstration) can clearly be determined empirically by an appropriate series of demonstrations and testings. The per cent of perfect performers drops from 100 to a very small value as length of demonstration increases, and the first D-A span is the point at which only 75 per cent can reproduce perfectly. Using this as a new starting point, the second D-A span can be determined in the same fashion as the next demonstration segment which can be retained well enough for 75 per cent of the learners to reproduce it perfectly. The entire task can thus be broken down empirically into its D-A span units, and if practice is interposed at each of these points, a relatively high level of correct practice is expected during each of the overt practice periods.

Whereas overt performance would be high, however, integration of the total task would be weakened by virtue of the fact that the learner (a) is practicing in fairly arbitrary units (how much he can "hold in his head" at once) and (b) is breaking up the transition between D-A span segments by taking time out to practice. Thus he may successfully practice a unit, but will he be able to put the units together? The issue is comparable to the classical "whole-versus-part" problem in which part learning is easier as far as parts are concerned but creates difficulties in fitting the parts together into an integrated whole.

The program of research reported in Chapters 3 and 4, and summarized in Chapter 5, is concerned with this basic issue. The research makes two kinds of comparisons after breaking the tasks down into empirical D-A spans: one is the whole-part type comparison in which practice can be interpolated after one D-A span, more than one D-A span, or at the end of the entire sequence; the other is the comparison of these procedures with a "transition" procedure in which single D-A span segments are used in the first demonstration-and-practice attempt, but in which larger segments are used in successive repetitions, culminating in a final uninterrupted demonstration and practice of the entire task. Another type of transition procedure was also used in which the learner was allowed to pace his own efforts on successive demonstration-and-practice sequences; that is, he was allowed to control the size of demonstration units before a time out for practice. Two quite different tasks were used: a mechanical assembly task in Chapter 3, and a geometric construction task in Chapter 4.

In Chapters 6 and 7, studies are reported in which attention is given to the concept of natural units of the task as optimum stopping points for overt practice. Also, in these studies, the demonstration material was constructed from the standpoint of maximizing the factors which would enhance the inherent organization of the task, which involved the serial learning of over 60 separate response elements. Part of the applied purpose, therefore, was to test the effectiveness of the demonstration and presumably of the principles used in preparing it. In Chapter 8, a particular method of enhancing the inherent organization of the sequential task is investigated, the main concept being that of the "perceptual blueprints" mentioned earlier.

Chapter 9 gives an overall summary and interpretation of the studies reviewed in Chapters 6, 7, and 8. It also presents, in concrete form, the principles used in constructing the demonstration. The "principles," of course, are for the most part in line with the theoretical considerations presented in the present chapter. The principles were not individually tested, but since the demonstration was highly successful, it reflects favorably on all of the principles used in its construction as well as on the theory which gave rise to the principles. Their status is that of "expert advice" based on theory and experience.

CHAPTER 3

OPTIMUM METHODS OF COMBINING PRACTICE WITH FILMED DEMONSTRATION IN TEACHING COMPLEX RESPONSE SEQUENCES: SERIAL LEARNING OF A MECHANICAL-ASSEMBLY TASK[1]

Garry J. Margolius
Fred D. Sheffield

Introduction

An inherent difficulty in the use of training films is that the learning which occurs while viewing the film is limited to acquisition of perceptual and symbolic responses which must later be translated into overt performance. The learner has a double problem. He must retain the visual and verbal contents of the film, and he must later utilize the new perceptual-symbolic material to guide his overt behavior. When the purpose of the training film is to teach a complex response sequence, such as assembly of a rifle, the training effectiveness of a single filmed demonstration may be fairly limited. The serial learning of the perceptual-symbolic sequence will be attenuated by intra-serial interference, and overt performance will be further attenuated by incomplete transfer—in the actual task situation—of the implicit responses acquired.

Presumably the above difficulties in the use of training films can be offset to some extent by interspersing overt practice periods in the course of a filmed demonstration. General considerations suggested the concept of a "demonstration-assimilation span" (D-A span), which refers to the amount of demonstration content which the learner can assimilate and translate into overt performance without much loss through interference from other parts of the demonstration and from the disparity between the film situation and the practice situation. General theoretical considerations further suggested the concept of "consolidation" of demonstration material, which here refers to the effects of an overt practice—at a time when assimilation of a segment of demonstration material is optimum—in preventing subsequent interference from assimilation of new portions of the demonstration content. It is presumed that practice at overt performance at the end of an appropriately short sequence will "consolidate" the sequence so that it combines by a cross-conditioning mechanism into an organized unit of elements, which elements hang together and are

[1]This chapter is based on a report first submitted on November 23, 1955, under contract AF 18(600)-1210 between Boston University and the U. S. Air Force. It was subsequently reproduced for limited circulation as a Laboratory Note for the Training Aids Research Laboratory (Margolius, Sheffield, and Robinson, 1956). See also footnote 1, Chapter 5.

less subject to interference from other elements, past or future. The argument is that even the same element, or a highly similar one, in a different part of the sequence, will interfere less because the individual element has become part of a perceptual pattern and no longer functions as a discrete stimulus or response. Finally, theoretical considerations suggested the concept of "natural units" of the task to be learned. Such units are a function of the particular serial-learning task and would not necessarily correspond with D-A span units. Elements of such units would be expected to be relatively free of intra-serial interference from elements in other units of a complex response sequence.

The purpose of the present research was to test the theoretical expectation that, in the joint use of filmed demonstration and overt practice, the appropriate use of practice periods interspersed through the film would be more effective than the use of practice only at the termination of the filmed demonstration. It was assumed in this research that overt practice adds significantly to the amount of learning obtained from the demonstration, and no control was used to illustrate the fact that demonstration plus practice is more effective than demonstration alone. Thus the main question was: given the use of both demonstration and practice, what is the most effective way of combining the two? The specific context in which this question was asked was that in which the task involved serial learning of a complex mechanical assembly which could not be acquired from a single filmed demonstration.

The demonstration-practice conditions considered in answering this question were the following:

1. D-A Span Segment practice. In this treatment the demonstration was stopped at the end of a given segment of the film and the learners practiced overtly what had been shown during the segment. The segments used had been previously empirically established as approximately one demonstration-assimilation span in length (see Chapter 2), although the subdivision was partly in terms of natural units.

2. Larger-Segment practice. In this treatment adjacent segments used in the first treatment were combined into larger segments and practice was interpolated at the end of each successive pair of segments used in the first treatment.

3. Whole practice. In this treatment the demonstration was continued through its entirety and practice was given only at the end of each complete showing.

4. Transition conditions. In this treatment the first showing utilized D-A span segments exactly as in the first treatment, the second showing utilized Larger-Segment practice exactly as in the second treatment, and the final showing utilized Whole practice exactly as in the third treatment. Thus transition was provided from initially smaller segments to the final stage of practicing the entire assembly as a whole.

5. Self-Pacing conditions. In this treatment the size of a demonstration-practice segment was left up to the individual learner. He was permitted to stop the demonstration when he felt he should practice what had been demonstrated thus

far, his instructions being, as in the case of all subjects, that after a set number of showings he would be tested on the entire assembly without help from the film.

The above five basic treatments were not studied in a single experiment; rather, two experiments were conducted, the first of which compared treatments 1 through 4 and the second of which compared treatments 4 and 5, together with a methodological variant of 4 to be described later. The two experiments will be presented separately.

Experiment 1

Purpose

The purpose of this experiment was to compare the first four treatments described above. Specifically, the purpose was to intersperse overt practice at the end of each "natural unit" and/or "demonstration-assimilation span" of demonstration material and compare the degree of learning with that produced with an equal amount of overt practice given (1) at the end of larger sub-units of the film, (2) at the end of the entire film, and (3) at the end of units of increasing size, starting with smallest and ending with the entire demonstration. In all four conditions the learners saw the entire demonstration three times and practiced the entire mechanical assembly three times; what was varied was the timing of demonstration and practice used prior to the final test.

Method

Demonstration and Practice Materials

A 16-mm. black and white film 18 minutes in length demonstrated the step-by-step assembly of an automobile ignition distributor.[2] The component parts were introduced in the film as needed, and the correct fitting together of these in a sequential order was depicted. An accompanying narration identified the parts and described the movements necessary to complete the assembly.

A standard automobile distributor of the same model as that depicted in the film demonstration was used for overt practice. The component parts, plus five confusion pieces, were laid out by the experimenter in a standardized pattern on a large white masonite sheet before each practice period. A screwdriver and tweezers were made available to the learner.

---

[2]This film was developed by J. J. Wulff in connection with contract AF 18(600)-35 between the United States Air Force and Yale University, for use in the study described in Chapter 10 of this volume, and was made available to Boston University by the Air Force Personnel and Training Research Center.

Subjects

Forty male undergraduate students at Boston University served as voluntary subjects. Each subject was assigned to one of four groups. All subjects were paid for the three-hour experimental period. Potential subjects were disqualified in cases of prior knowledge about the automobile distributor.

Design

All subjects viewed the entire film three times and were given an opportunity to practice the assembly in conjunction with each demonstration. At the conclusion of training each subject was asked to assemble the distributor without any further film instruction. This fourth and final assembly was considered a terminal performance test.

As already noted, the critical experimental differences among subjects were concerned with the method of interpolating demonstration and practice portions of training. Preliminary study had indicated that the film could be divided into four sections, each roughly approximating a demonstration-assimilation span. The divisions were made in terms of ability to assemble the actual distributor parts after viewing a segment of the film rather than in terms of film footage. The first and last of the four sections selected were judged to be "natural units" of material. The middle two sections together were judged to comprise a "natural unit," but a unit too long and complex to be assimilated in a continuous showing. This middle "natural unit" was therefore subdivided into two D-A span units. The film divisions thus chosen served as the focal points around which the experimental manipulations could occur during training. The four training procedures were differentiated with respect to the number of film sections viewed by the learner before training practice was attempted during each of the three training sessions. The training schedules are outlined in Table 1, where D and P stand for demonstration and practice respectively, and the numerals refer to the four specific sections within the training task.

Details of Training Procedure

D-A Segment Method. The subject was seated at a table upon which the distributor parts had previously been placed and covered from view. He was instructed to observe and listed carefully to the material presented in the film since he would soon be called upon to demonstrate his knowledge. The room light was then lowered and the film projector started. For this group the projector was stopped at the end of D1 and the room light was turned up. The subject was then instructed to select the parts necessary to complete the assembly demonstrated thus far, as quickly and accurately as possible. He was reminded of the name of the last piece necessary and the cover was removed from the distributor parts. When the subject indicated that he was finished with his selection, the experimenter recorded both time and errors, corrected the selection if necessary, and then covered the remaining, or non-selected, parts. The subject was further instructed to assemble the correct

TABLE 1

Conditions Used in Experiment 1

Group	Trial	Demonstration-Practice Sequence
I. D-A Span Segments	1	D1, P1; D2, P2; D3, P3; D4, P4
	2	(Repeat)
	3	(Repeat)
II. Larger Segments	1	D12, P12; D34, P34
	2	(Repeat)
	3	(Repeat)
III. Whole Task	1	D1234, P1234
	2	(Repeat)
	3	(Repeat)
IV. Transition	1	D1, P1; D2, P2; D3, P3; D4, P4
	2	D12, P12; D34, P34
	3	D1234, P1234

selection of parts as quickly and as accurately as possible and to notify the experimenter upon completion. While the subject was at work his temporary errors—those he made but corrected during the practice period—were recorded by the experimenter. When the subject indicated completion or exceeded the time limit (Table 2), the practice material was removed and a record was made of fixed (uncorrected) errors and assembly time. Each uncorrected faulty response or omission of a part counted as a fixed error. Assembly-time score did not include time taken to screw parts together. The experimenter stopped his watch whenever the subject placed the screwdriver in position and started it again when the subject removed the screwdriver, in order to minimize differences due to mechanical aptitude.

Sections 2, 3, and 4 were run in a similar fashion. Before each period of assembly practice, the experimenter corrected any fixed errors in the previous attempt. For example, when the subject was ready to begin the assembly practice of Section 3, he was handed a correct assembly through Sections 1 and 2. He never observed corrections being made, however; he was simply informed at the beginning of the next practice unit that the assembly up to that point was now correct. Following the four film-and-practice periods, the entire procedure, as shown in Table 1, was repeated a second and then a third time. Table 2 indicates film length, time limits for selection and assembly, and number of correct responses for each of the four sections used in the D-A Segment group.

TABLE 2

Timing of Film Footage and Limits Set on Practice Time for the Four Sections of the Demonstration[3]

Section	Film Time		Practice Time: Selection		Practice Time: Assembly		Maximum Possible Correct Responses
	Min.	Sec.	Min.	Sec.	Min.	Sec.	
1	5	30	2	15	4	30	9
2	2	30	1	15	3	00	4
3	4	00	1	15	5	00	7
4	6	00	2	15	7	30	10
Totals	18	00	7	00	20	00	30

Larger-Segment Method. This group viewed Sections 1 and 2 of the film as a single unit, followed by practice of both sections; they then viewed Sections 3 and 4 as a single unit, followed by practice of both these latter units. In practicing Sections 1 and 2 the subjects were scored for P1 and P2 separately, and any mistakes in assembly at the end of P1 were corrected before starting P2. The same procedure was used when they practiced P3 and P4. Thus practice was made identical in form to that used in the D-A Segment method, the difference being that the Larger-Segment group practiced two smaller segments in immediate succession rather than with an interpolated film section. The purpose of this procedure was to make practice (and measurement) conditions identical for both groups, except for the size of the demonstration-practice segment involved. The procedure as described was repeated three times prior to the final test.

Whole Method. This group viewed the entire film without interruption as a single demonstration unit. They then practiced the entire assembly but, as in the first two groups, they were scored separately for P1, P2, P3, and P4, and any necessary corrections were made at the end of each of these practice sections. This arrangement of demonstration and practice was repeated as before for a total of three complete demonstration-and-practice periods before the final test.

Transition Method. This group was also given three complete demonstration-and-practice periods, but these were arranged so that the first period was identical with that of the D-A Segment method, the second period was identical with that of the Larger Segment method, and the final period was identical with that of the Whole method.

[3]Time limits for both total task and sections of the task were assigned on the basis of pre-test results so that all subjects with minimum knowledge necessary for task completion could easily finish both selection and assembly within the times allotted.

Final Test Procedure

For the final test, the subject was once again presented with the distributor parts and instructed to assemble the complete mechanism without preliminary selection of parts and without any break between units. The experimenter recorded assembly time and errors for each of the four sections separately as before, although the subject proceeded without a break and without any corrections. It should be noted that while the total assembly time for the final test could be determined accurately, the part-times for the separate units were in some cases approximate because of variations introduced by the subjects in the order of assembling the parts.

Results

The present analysis will first treat performance during the training trials, and then performance during the final test. As selection opportunities during the final test could not be made comparable to selection during training, the training-selection data are omitted here.

Performance During Training

The mean number of fixed errors[4] made by each group during all training trials is given in Table 3 and presented graphically by successive trials in Figure 1. The summary of an analysis of variance applied to the error data summed for all practice sessions is given in Table 4. F, for differences among the four groups, attained significance at the .01 level of confidence. The results of comparisons between means are given in Table 5. All P values are for one tail. Significance ratios reported for all comparisons except that between smaller segments and transition attain significance at probability values varying from .01 to .07.

TABLE 3

Means and σ^2 of Total Fixed Errors During Practice

Group	Mean	σ^2
D-A Segment	6.5	26.05
Larger Segment	11.1	32.89
Whole	18.4	115.60
Transition	6.7	26.21

[4]It will be recalled that a "fixed" error is defined to include (a) an incorrectly assembled part and (b) a part left over because the subject did not know how to include it.

TABLE 4

Summary of Analysis of Variance
for Total Fixed-Error Scores During Practice

Source	df	Mean Square	F
Between Groups	3	343.63	6.48*
Within Groups	36	53.00	

*Significant at the .01 level

TABLE 5

Comparisons Between Means of Fixed-Error Scores During Practice

Comparisons	t	df	p*
Whole—Transition	2.99	18	<.01
Whole—D-A Segment	3.01	18	<.01
Whole—Larger Segment	1.80	18	<.05
Larger Segment—Transition	1.72	18	<.07
Larger Segment—D-A Segment	1.81	18	<.05
Transition—D-A Segment	.08	18	---

*p values are given for one tail

A similar analysis was undertaken for total-error score, consisting of fixed-plus-temporary errors. Means, analysis of variance, and group comparisons are given in Tables 6, 7, and 8, respectively. Once again all comparisons except smaller-segment versus transition attain appreciable levels of significance, with P values varying from .01 to .05.

Time measures taken during each training session are shown in Figure 2. The similarity of these data to the error data reported is readily apparent. It was felt that the two measures taken, time and correct responses (total possible correct minus fixed errors), might advantageously be combined to give a single meaningful index of performance efficiency. The index—number of correct responses per unit of time—was obtained by dividing total correct responses by assembly time. The ratio so obtained, called "performance rate," was computed for each subject at each practice session. The results with this performance-rate measure are shown graphically in Figure 3.

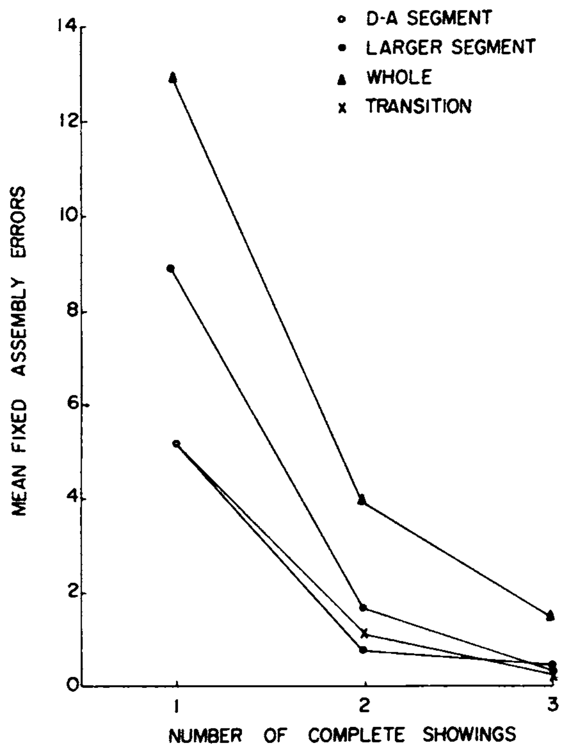


Figure 1. Mean Fixed Errors After Successive Film Showings, Presented as a Function of Size of Demonstration-Practice Segment Used During Training.

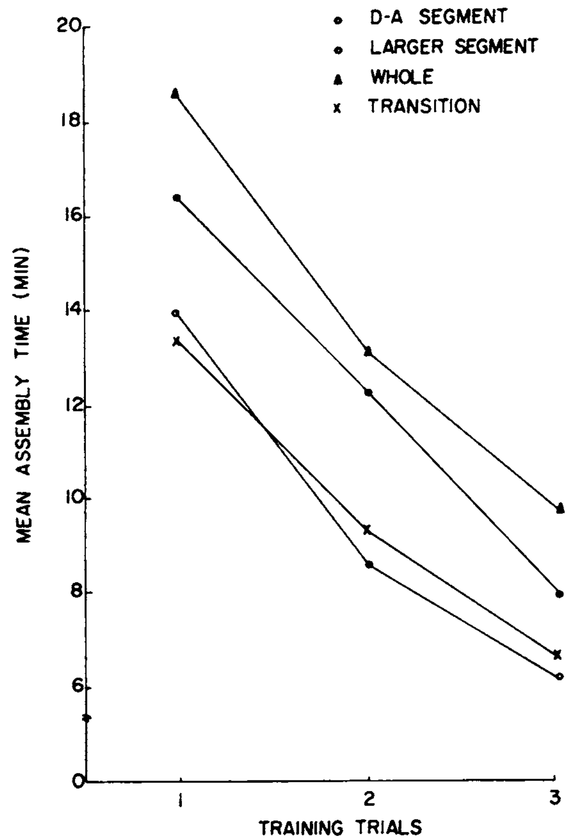


Figure 2. Mean Assembly Time After Successive Film Showings, Presented as a Function of Size of Demonstration-Practice Segment Used During Training.

TABLE 6

Means and Variances of Combined Fixed and Temporary-Error Scores During Practice

Group	Mean	σ^2
D-A Segment	11.0	45.60
Larger Segment	16.7	51.21
Whole	24.0	93.00
Transition	10.6	36.24

TABLE 7

Summary of Analysis of Variance for Combined Temporary and Fixed-Error Scores During Practice

Source	df	Mean Square	F
Between Groups	3	393.09	11.22*
Within Groups	36	35.02	

*Significant at the .01 level

TABLE 8

Comparisons Between Means of Combined Fixed and Temporary-Error Scores During Practice

Comparisons	t	p*
Whole—Transition	3.68	<.01
Whole—D-A Segment	3.40	<.01
Whole—Larger Segment	2.31	<.05
Larger Segment—Transition	1.96	<.05
Larger Segment—D-A Segment	1.74	<.05
Transition—D-A Segment	0.13	---

*p values are given for one tail

All of the increases shown in the figure at successive stages of practice are significant at better than the .01 level in each of the four groups, the _t_-ratios involved in the gains ranging from 3.73 to 12.40. It will be noted, however, that the acquisition curves tend to diverge slightly, performance rates being further apart after the third demonstration-practice period than after the first. To test for this apparent interaction effect the <u>Groups x Practice</u> variance was compared with the <u>Subjects x Practice</u> variance. The apparent interaction fell just short of significance at the .05 level, suggesting that there may be a real difference in the direction of divergence of performance-rate measures, the initially superior training methods becoming more superior with increased training.

The above results are shown in Table 9, along with a summary of the analysis of variance involving Group effects and Practice effects. The F for Groups was determined by using Subject variance as the error term; the F for Practice was determined by using Groups x Practice as the error term. As can be seen, the effects of experimental treatment (Groups) and practice are significant beyond the .01 level.

Individual means and variances are shown in Table 10, and the results of significance tests of group-mean comparisons at each stage of training are given in Table 11. Aside from the essential equality of the D-A span and Transition groups, all comparisons are significant at the five per cent level or better, except that between the Whole and Larger-Segment methods for the second trial.

Final Test

Turning to the final test it can be noted that five of 10 individuals in the whole group completed an errorless performance. When this proportion is compared with the performance of other groups combined (27/30), the difference attains significance at the .01 level (one tail). Although this crude measure of proficiency clearly indicates the inferiority of the Whole method versus others, a more sensitive index, performance rate, was utilized in an effort to discern inter-group differences.

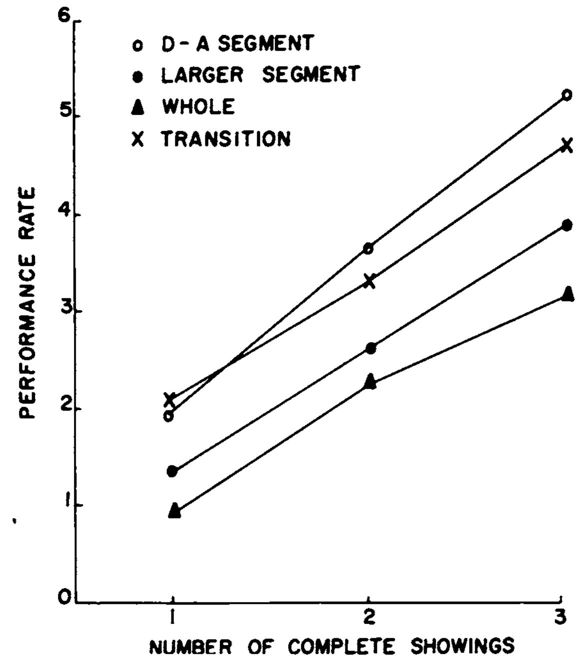


Figure 3. Performance Rates at Successive Stages of Practice.

Means and Variances of Performance Rates During Final Test are given in Table 12. An analysis of variance applied to performance rates during the final test is

TABLE 9

Summary of Analysis of Variance of Performance Rates at Successive Stages of Practice

Source	df	Mean Square	F
Groups	3	14.15	6.86*
Practice	2	'68.58	131.9*
Subjects	36	2.05	
Groups x Practice	6	0.52	2.17
Subject x Practice	72	0.24	

*Significant beyond the .01 level

TABLE 10

Means and Variance of Performance Rates at Successive Stages of Practice

Group	Practice 1 Mean	Practice 1 σ^2	Practice 2 Mean	Practice 2 σ^2	Practice 3 Mean	Practice 3 σ^2
D-A Segment	1.96	0.57	3.64	0.68	5,17	1.68
Larger Segment	1.38	0.28	2.59	0.88	3.84	0.75
Whole	0.94	0.13	2.21	0.72	3.14	0.67
Transition	2.10	0.81	3.27	0.65	4.69	1.29

TABLE 11

Performance Rate Comparisons Between Group Means at Successive Stages of Practice

Comparisons	Practice 1 t	Practice 1 p*	Practice 2 t	Practice 2 p*	Practice 3 t	Practice 3 p*
Whole—Transition	3.60	<.01	2.72	<.01	3.31	<.01
Whole—D-A Segment	3.67	<.01	3.63	<.01	3.97	<.01
Whole—Larger Segment	2.87	<.03	0.91	<.20	1.77	<.05
Larger Segment—Transition	2.07	<.03	1.65	<.07	1.78	<.05
Larger Segment—D-A Segment	1.91	<.05	2.52	<.01	2.55	<.01
D-A Segment—Transition	0.35	---	0.96	---	0.83	---

*All p values are given for one tail.

TABLE 12

Means and Variances of Performance Rates During Final Test

Group	Mean	σ^2
D-A Segment	4.91	0.75
Larger Segment	4.02	0.65
Whole	3.22	0.72
Transition	4.95	0.79

summarized in Table 13. Comparisons between group means are shown in Table 14. Results of the tests indicate essential agreement with those obtained for the practice periods. Thus the groups having demonstration and practice, either in D-A span segments or having transition from these segments to the whole task, performed about equally well in the final test, but both of these conditions were superior to the practice method involving larger than D-A span segments, and all segmental methods were superior to the Whole method.

TABLE 13

Summary of Analysis of Variance
of Performance Rates During Final Test

Source	df	Mean Square	F
Groups	3	6.77	8.93*
Within Cells	36	0.78	

*Significant at the .01 level

TABLE 14

Comparisons Between Performance-Rate Means
of all Groups During Final Test

Comparisons	t	p*
Whole—Transition	4.23	<.01
Whole—D-A Segment	4.18	<.01
Whole—Larger Segment	2.11	<.03
Larger Segment—Transition	2.32	<.03
Larger Segment—D-A Segment	2.28	<.03
D-A Segment—Transition	0.10	---

*p values are given for one tail

Experiment 2

Purpose

A major purpose of Experiment 2 was to obtain information on Condition V listed in the Introduction section, namely, the Self-Pacing condition, in which the learner has control over the length of the demonstration-practice segments.

The relative advantages and disadvantages of a self-pacing method are not very predictable, partly because judgment as to optimum conditions of learning are left up to the learner. Two clear-cut theoretical factors should be mentioned: (1) the factor of individual differences in D-A span and (2) the dependence of natural units on the nature of the task.

The first of these factors—individual differences in D-A span—follows from normal expectation of individual differences in ability to assimilate and remember and implies a possible gain from self-pacing. Thus individuals with differing ability to "hold in their heads" what is shown in the demonstration will have different lengths of D-A span; to the extent that they can be trusted to stop the film when they reach a "saturation point," they should benefit from a procedure in which they have control of the points at which they take "time out" to consolidate what they have seen in the demonstration with overt practice. Considering only this factor, i.e., ignoring natural units, an advantage to differential pacing would be a definite prediction, although the question still remains as to whether with self-pacing the learner will stop the film when he actually has reached the "saturation point." Alternative possibilities are that the learner will be over-cautious and stop the film before he needs to or will be over-confident and let the film run past his personal D-A span.

The second factor—that of dependence of natural units on the nature of the task—implies a loss from self-pacing, the assumption being that natural units depend on the task regardless of learning ability, and that the individual learner would be a poor judge of where to subdivide the task along lines of "natural units" for learning. Theoretical considerations suggest that, except for long or complex natural units, it is better to introduce consolidation through practice at the end of natural units of the task, even if these are somewhat more than or less than a D-A span unit. The learner is expected to integrate the task eventually along lines of task-determined natural units, but at the outset of training he would not be expected to anticipate or recognize these in his control of practice methods. Thus the factor of natural units suggests that the learner would be better off if practice periods were imposed at certain points in the task than he would be if he selected his own practice periods.

The considerations above bring out the fact that the relative success of the Self-Pacing method depends to a large extent on the judgment of the learner. For this reason the variable of self-pacing is in a special category as an experimental variable. The experimenter can manipulate the degree of control permitted the subject, but he does not manipulate what the subject will expose himself to in the way of causal factors. Without a large N, the latter empirical question cannot be given a definite answer, and the present experiment on the self-pacing variable is regarded as a pre-test or pilot experiment indicating the range of possible outcomes rather than a definitive experiment on the effects of a particular experimental manipulation.

A second purpose of Experiment 2 was to assess the factor of assembly correction used in the practice procedures of Experiment 1 in the case of the Larger Segment method, Whole method, and Transition method. Because of the interest in obtaining comparable measures from all groups during the practice sessions of Experiment 1, each section of the task was corrected before the next section was attempted. For the Whole group, this meant that although the film was viewed in its entirety, the practice following demonstration might well be termed "corrected Whole method." Not only does this procedure reinstate important cues before each new section is undertaken, but also it minimizes the compounding of errors which probably would accompany a Whole-practice period in which errors were allowed to accumulate. It was felt that these factors raised levels of performance in the practice and final tests of not only the Whole-method group and the Larger-Segment group, but also the Transition group, in which correction was used in the second and third practice sessions.

On the basis of pre-testing it was felt that a demonstration of the advantage of the section-correction method of practice under whole-demonstration conditions would be trivial. Documentation of the reasons for this conclusion can be obtained by referring back to Figure 1 (of the Results section of Experiment 1). Even with the "corrected Whole method" the fixed errors averaged over 40 per cent during the first practice with the Whole method used; the expectation was that if fixed errors were allowed to accumulate the subjects would have been essentially unable to practice the final section of the demonstration.

A more realistic and instructive condition for determining the effects of the correction method of practice is provided by the Transition condition of Experiment 1. For one thing the Transition method appears to be one of the preferred training procedures on the basis of Experiment 1, giving some practical significance to the question of whether section correction during practice produces better learning. For another thing an important empirical question is involved in the case of the Transition method since two theoretical factors work in opposite directions when section correction is used. On the one hand, the correction procedure increases the likelihood that the learner will be able to practice and consolidate the correct responses during the overt-practice periods. On the other hand, however, the Correction method is somewhat contrary to the integration principles involved in the Transition method. Thus the method involves an attempt to integrate the sub-units of the task into larger and larger superordinate units as training progresses. To a certain extent, therefore, this purpose is defeated by taking time out for a new selection of parts before continuing with the corrected assembly.

The choice of Transition conditions as the setting in which to investigate the correction factor in practice was also influenced by the expectation that the Self-Pacing group would probably start with small segments of demonstration and practice and work up to larger ones with successive showings. Thus the condition in Experiment 1 which was expected to be most comparable to self-paced efforts was condition IV, involving corrected practice and the Transition method. It was felt necessary to repeat exactly at least one of the conditions of Experiment 1 in order to provide an anchoring point for assessing the Self-Pacing group since the two experiments were done at different times and involved somewhat different subject populations.

Method

Subjects

The subjects were 30 volunteer undergraduate male students attending the summer session at Boston University. The students were paid for the three-hour experimental session.

Design

In line with the foregoing purposes the subjects were divided into three groups according to experimental treatment: a Transition group exactly as in Experiment 1, a Transition group differing in the omission of section correction during practice, and a Self-Pacing group in which the learner made his own decision as to when to stop the film for a practice session. The details of design and method follow. Specifically the training methods investigated were:

Group 1. Transition method with section correction during practice. This method will be referred to as the T_1 method; the treatment was identical with the Transition method (Group IV) used in Experiment 1, except for a minor change in initial instructions.

Group 2. Transition method with practice units corresponding to demonstration units. This method will be referred to as the T_2 method. As an example of the difference between T_2 and T_1, during the second film showing where the T_2 subject has viewed sections 1 and 2 of the film as a unit, he also practiced them as a unit without interruption. Correction of the assembly was not made following practice of section 1, as in the T_1 method, but was withheld until an attempt at both sections was completed. Similarly, after the third showing there were no breaks or corrections at all in practice following the film demonstration.

Group 3. Self-Paced method. For each of the three demonstrations the experimenter stopped the film whenever the subject signalled his readiness to practice as much of the assembly as he had just seen demonstrated. This procedure was continued until the entire demonstration had been viewed and practiced. Thus, for this group (S-P) the number of practice units depended upon the subject. Time limits were set for every possible step choice in the demonstration, and thus for any practice unit selected by the subject. Total time of any combination of sections selected was equal to that set for all other groups. As was the case for T_1, assembly errors were corrected before S went on to the next practice unit.

Except for the above-noted variations in treatment, all conditions of Experiment 2, both general and specific, in the training and testing procedures were identical with those employed in Experiment 1.

Results

As in Experiment 1, the time scores and the fixed-error scores showed essentially the same pattern of results. The results will therefore be presented in terms of performance rate, which used both time and fixed errors in a joint measure of performance per unit time. Performance rates during training in Experiment 2 are shown in Figure 4, along with the Transition-method data of Experiment 1. The means and variances for the data from Experiment 2 plotted in the figure are shown in Table 15. Comparable data for the final test are shown in Table 16.

It will be recalled that partly for anchoring purposes the T_1 treatment in the present experiment was run under conditions essentially identical with those of the Transition method of Experiment 1. It can be seen in the figure that the acquisition results are quite comparable for the two experiments under these conditions.

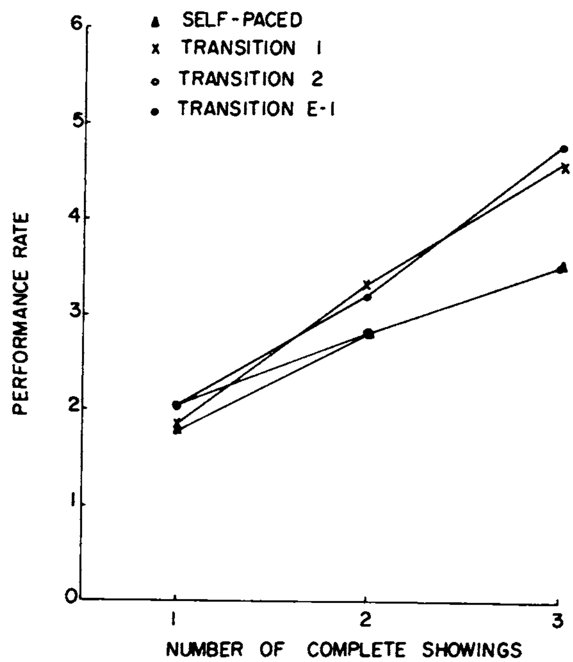


Figure 4. Performance Rates at Successive Stages of Practice for Groups in Experiment 2 and for the Transition Group in Experiment 1.

Comparison Between T_1 and S-P

In line with the first purpose of Experiment 2 the results for the Self-Pacing condition were compared with those for T_1. This comparison permits an assessment of the S-P method in the same setting as that for the first experiment, in which T_1 and the D-A span segments were essentially equally effective. In making this comparison, note in Figure 4 that the two groups start out at the same level but diverge with practice, the T_1 group being superior by the end of the third showing. To test the significance of this divergence, gain scores were utilized, each person's gain score being the difference between his performance rates for the third and the first practice. The mean gain can be obtained from Table 15 as (4.63-1.88) = 2.75 for T_1 and (3.53-1.81) = 1.72 for S-P. The second-order difference of 1.03 in performance-rate increase has a *t* of 2.17, which for the 18 df involved is significant at the .05 level (both tails).

Turning to the relative performances of T_1 and S-P in the final test, the difference shown in Table 16 of (4.48-3.35) = 1.13 has a *t* of 1.57 with P value of .20 (both tails).

TABLE 15

Means and Variances of Performance Rates
at Successive Stages of Practice

Group	Practice 1		Practice 2		Practice 3	
	Mean	2	Mean	2	Mean	2
Transition 1	1.88	0.46	3.38	0.77	4.63	1.35
Transition 2	2.08	0.62	2.81	1.28	3.50	2.53
S-P	1.81	0.17	2.84	1.02	3.53	1.41

TABLE 16

Means and Variances of Performance Rates
During Final Test

Group	Mean	2
Transition 1	4.48	1.26
Transition 2	4.24	2.68
S-P	3.35	3.49

Thus, both during training and in the final test, the T_1 method was superior to the S-P method, although the superiority was not very reliable statistically. The statistical significance of the differences is increased, however, if T_1 results are combined with the Transition-method data from Experiment 1. This combination is justified on the grounds that the treatments were essentially identical in the two experiments and differences in results between them were not at all significant. With this combination the gains during training of the section-correction Transition method were superior to the Self-Pacing method with $t = 2.44$, and $P = <.05$ (both tails). Using the same combination this Transition method was superior in the final test with $t = 2.47$, and $P = .02$ (both tails).

Thus it appears that this Transition method was superior to the Self-Pacing method, given the conditions of self-pacing used in this experiment. It should be noted that these conditions were not necessarily optimum for self-pacing in general. Because of the exploratory nature of this part of the research, the learners were given considerable freedom in their method of pacing themselves. It is possible that with special instructions on appropriate methods of self-pacing this procedure would be more effective.

It is of some interest to examine the S-P group from the standpoint of actual number of stopping points used. As was expected, progressively fewer stopping points were used with successive showings of the film. At all showings, however, a larger number of stops was used by the S-P subjects than the 4-2-1 progression imposed in the Transition groups. The relevant data are shown in Table 17. Part of the apparent inferiority of the S-P group in performance rate may be due to the use of too many stops. In this connection it is interesting that there was a significant (P = .02, both tails) negative correlation (-.70) between the number of stops made by the subjects during the demonstration and their performance on the final test. This correlation is in line with the expectation that too many stops would tend to decrease the amount learned, but the correlation is contaminated with the likely inverse relation between learning ability and number of stops. Thus the poor learner would tend to stop often in order to be able to practice what he had seen before the demonstration got beyond his memory span.

TABLE 17

Summary of Number of Units Chosen
for Each Training Session in Self-Pacing Procedure

	1st Showing	2nd Showing	3rd Showing
Median	6.5	3.5	2.0
Mean	6.9	4.6	2.8
σ^2	17.7	9.6	6.3

Comparison Between T_1 and T_2

In line with the second purpose of Experiment 2, the T_1 and T_2 groups were compared by essentially the same tests as were used in the T_1 versus S-P comparisons just described. As can be seen in Figure 4, the T_1 and T_2 groups start close together after the first showing (in which demonstration and practice is identical for both groups) but diverge in the second and third showings. Using overall gain scores (third-minus-first practice performance rate) the mean gain for T_1 is (4.63-1.88) = 2.75 whereas that for T_2 is only (3.50-2.08) = 1.42. The $\underline{t}$ for this difference is 2.68, with P value of .01 (one tail) in the expected direction. The corresponding figures when T_1 is combined with its identical counterpart in Experiment 1 (i.e., Group IV) are $\underline{t}$ = 3.10, with P value of .01 (one tail) in the expected direction. Thus it appears relatively safe to conclude that the T_1 method of Experiment 2 (Transition method of Experiment 1) produces better performance during practice than T_2. This is not a surprising result in view of the advantages during practice provided by the section correction in T_1.

A perhaps more surprising result is the lack of much difference between T_1 and T_2 in the final test performance. The means of 4.48 for T_1 and 4.24 for T_2 do not differ at all significantly (t = 0.36, P = .80 in the expected direction), and even if T_1 is combined with the corresponding group in Experiment 1, the combined mean of 4.72 does not differ significantly from T_2 (t = 0.93, P = .40 in the expected direction). These results suggest a definite positive transfer effect from the absence of section correction in the T_2 treatment. That is, practice and test conditions were more alike with T_2, especially in the practice after the third showing, which corresponded exactly with final test conditions. Evidence for this factor can also be seen in the fact that, whereas T_1 dropped from a performance rate of 4.63 in the third practice to 4.48 in the final test, T_2 rose from a performance rate of 3.50 in the third practice to a rate of 4.24 in the final test. The significance of this differential shift was tested using each subject as his own control (i.e., final-test rate minus third-practice rate per individual subject). The t for the second-order difference was 2.26, giving a P-value of .05 (both tails).

Interpretation

If one reviews the data for final test performance in both experiments, it can be seen that the poorest result in terms of performance rate was found with the Whole method (performance rate = 3.22) of practice used in Experiment 1. All of the other methods (six groups in all) involved some form of practice interspersed through the demonstration. At the most gross level of conclusion, therefore, it would appear that interspersed practice periods can be used to advantage in training films demonstrating complex mechanical assemblies. It should be recalled that time spent in practice was identical for all groups. Thus the above conclusion does not merely imply that interspersed practice adds to what the film teaches; rather it implies that practice during the film can be made superior to practice at the end of the film. The conclusion applies, of course, only to equally lengthy and complex assembly tasks.

Of all of the different methods of interspersing practice during the demonstration, the Self-Paced method (Experiment 2) gave the poorest performance in the final test (Mean = 3.35). The generality of this result must be questioned, however, if only because of the possibility that the summer-session students were an inferior population compared to the regular students used in Experiment 1. Thus the riskiest comparison is between the Self-Paced method of Experiment 2 and the Larger Segment method of Experiment 1 (Mean = 4.02). Beyond this question of generality due to sampling possibilities, the apparent inferiority of the Self-Paced method is of dubious generality because of the large amount of freedom permitted the learners in choice of demonstration and practice intervals. Thus the exact wording of the instructions could play a large role in the method adopted by the learner, and different results of the Self-Paced method might have been obtained if the subjects had been told, for example, to make their segments as long as possible, to try to increase the length with successive showings, and to try to select natural stopping points even if this meant making a unit a little longer than they could remember perfectly the first time. It seems unlikely that the wide range of intervals selected by the learners represents equivalent variation in learning ability, and this range could no doubt be reduced by appropriate instructions even within the framework of self-pacing.

The preferred methods of interspersing practice appear to be either the use of D-A span segments of demonstration and practice or the use of a Transition method starting with such units and progressing to successively larger units. Theoretically some form of transition would provide the most effective method of integrating the entire task. In the present research, however, the slight superiority of the Transition method over the D-A span segments on the final test (Experiment 1) was not at all significant. This may reflect too rapid transition in the Transition method used. Thus the advantage of consolidation of material is lost if the transition shifts the training procedure too rapidly in the direction of the Whole method of demonstration and practice. Transition should keep pace with consolidation or the gains will be cancelled rather than being additive.

An alternative possible explanation of the fact that the Transition method did not prove to be superior to the D-A span method is that in the progression to larger segments the larger segments were not combined along lines of natural units. Thus the film had four D-A spans but was judged to have only three natural units, the middle one being two D-A spans in length. In the Transition method, the larger segments used consisted of (a) the first natural unit plus half of the second one (these comprising the first two D-A spans), and (b) the second half of the second natural unit plus the last natural unit (these comprising the last two D-A spans). It may well be that a transition which merely combined the middle two D-A spans into the middle natural unit would have proven a more effective procedure than any of those used in the two experiments.

The results for the comparison of T_1 and T_2 in Experiment 2 make the point that without some sort of correction of errors during practice the learner's performance will suffer during practice. If the extreme of using no section correction with the Whole method had been tried, there would no doubt have been a huge difference both in practice and final test compared with the Whole method of Experiment 1. Such a group was not utilized because it was considered that the results would be trivial. The omission of section correction in the Transition method is less serious because the first practice is by section and therefore automatically involved section correction in the first demonstration and practice session. The T_1 versus T_2 comparison is most notable, however, for the lack of much loss in final test due to absence of section correction during practice periods after the first. This failure to get much difference is interpretable as being due to the advantage of training under conditions which simulate actual test conditions. To the extent that this interpretation is correct, the T_1 versus T_2 comparison actually bolsters the argument in favor of transition. It also bolsters the argument in favor of practicing along lines of natural units and providing an opportunity to integrate the elements within these. Thus, in the third practice the T_2 subjects were not interrupted by separate section performance (especially in the middle natural unit), and apparently as a result showed an improvement from third-practice to final-test performance.

CHAPTER 4

COMBINING PRACTICE WITH DEMONSTRATION IN TEACHING COMPLEX SEQUENCES: SERIAL LEARNING OF A GEOMETRIC-CONSTRUCTION TASK[1].

Walter Weiss
Nathan Maccoby
Fred D. Sheffield

Introduction

The preceding chapter reports research findings on the differential effects of several alternative methods of combining overt practice with a filmed demonstration in teaching a complex mechanical-assembly task. The main purpose of that study was to test the theoretical expectation that intra-serial interference, arising in the attempt to learn a complex serial task from a filmed demonstration, could be offset to a certain extent if overt-practice periods were interspersed at strategic points during the demonstration. The comparison condition in the above study was an equal amount of practice given only at the end of the demonstration. This training procedure gave the poorest performance of the several alternative methods employed.

The present research was essentially a parallel study having the same purpose but using a geometric-construction task rather than a mechanical-assembly task. A geometric-construction task (using pencil, paper, ruler, and compass) differs in several ways from a mechanical-assembly task, but probably the two most basic distinguishing features are: (1) the geometry task, being two-dimentional, can be viewed in its entirety at all stages of construction; and (2) the geometry task, being construction rather than assembly, is strictly a recall task involving reproduction from memory without benefit of external cues except those supplied by the learner himself. Thus, in a typical mechanical assembly, different parts get covered from view as assembly proceeds to completion; whereas in geometric construction all parts of the "assembly" are visible even upon completion. Also, in the assembly task, all of the parts to be assembled are present at the outset to provide cues for

[1]This chapter is based on a report first submitted on November 23, 1955, under contract AF 18(600)-1210 between Boston University and the U. S. Air Force. It was subsequently reproduced for limited circulation as a Laboratory Note for the Training Aids Research Laboratory (Weiss, Sheffield, & DuBois, 1956). See also Footnote 1, Chapter 5.

their manipulation, making the memory task partly that of recognition and reconstruction (memory for previous arrangement) rather than the pure recall involved in geometric construction.

The demonstration-practice conditions investigated in the present study were the following:

1. D-A Span Segment practice. In this treatment the filmed demonstration was stopped after a given segment of the film, and the learners practiced what had just been shown during the particular segment. The lengths of these segments were determined in a pre-test to approximate one "demonstration-assimilation span" (D-A span) each, defined for the purposes of this study in terms of the amount of demonstration material which could be assimilated and translated immediately into perfect overt performance by more than 75 per cent but less than 100 per cent of subjects used in the pre-test sample.

2. Larger-Segment practice. In this treatment, adjacent D-A span segments used in the first treatment were combined into larger segments (more than one D-A span). There were two sub-treatments of Larger Segment practice: (a) a treatment in which the larger segments were two D-A spans in length and (b) a treatment in which the larger segments were three D-A spans in length.

3. Whole practice. In this treatment the demonstration was continued through its entirety and practice was given only at the end of each complete showing.

4. Transition condition. In this treatment the first demonstration and practice utilized "natural units" of the construction task (which did not correspond exactly with D-A span segments); the second demonstration and practice combined adjacent natural units; and the third (final) showing utilized the Whole method, as in treatment 3.

5. Self-Pacing conditions. In this treatment the size of the demonstration-practice segment was left up to the individual learner. He was permitted to stop the filmed demonstration whenever he felt that an interpolated practice of what he had seen was advisable, his instructions being that after a fixed number of showings he would be tested over the entire task without help from the film.

6. "Natural-Units" practice. In this treatment, identical with the first showing of the transition conditions above, the demonstration and practice segments in all trials were those judged by the experimenters to be "natural units" (see Ch. 2, p. 29) from the standpoint of integration of sub-elements.

The foregoing treatments are similar in general to the corresponding treatments in the preceding chapter using the mechanical assembly task. For practical reasons and because of inherent differences in the tasks, however, the treatments can not be regarded as identical. Thus the present experiments should not be regarded as an exact repetition of the previous experiments, differing only in the use of a different task.

The present study was carried out as two separate experiments, the first being conducted with a group procedure in high school classes and the second being conducted with an individual procedure among college students. Treatments 1, 2a, 2b, and 3 were compared using the former procedure, and Treatments 3, 4, 5, and 6 were compared using the latter procedure. The two experiments will be presented separately.

Experiment 1

Purpose

The purpose of the experiment was to compare the first three treatments. Specifically, the purpose was to intersperse practice at the end of each "demonstration-assimilation span" of demonstration material, and to compare the degree of learning with that produced by an equal amount of overt practice given at the end of longer sub-units of the film or given only at the end of the entire film. Since Treatment 2 utilized two sub-treatments, the purpose actually involved comparisons of four different sets of training conditions. In all of these conditions the learners saw the entire demonstration twice and practiced the entire geometric construction twice before the final test. The nature and size of the demonstration-practice segment used were varied.

Method

Demonstration and Practice Materials

A film was shown demonstrating the construction of a five-sided equilateral polygon, using pencil and paper and a straight-edge and compass.[2] The starting point and the completed construction are shown in Figure 1. The film depicted this construction in step-by-step close-up shots, with narration describing each successive step. The task involved nine distinct operations or steps but was treated as having only six D-A span segments. That is, in three cases two adjacent steps were treated as comprising together a single D-A span segment. This was justified partly on the basis of a pre-test on 42 Boston University undergraduates. In this pre-test all but one of the six segments above could be performed perfectly by at least 75 per cent of the group immediately after the demonstration of the segment. The one segment which fell short of this criterion of a D-A span was a combination of the fifth and sixth steps, and was performed perfectly by 71 per cent of the pre-test sample. Since this performance was close to the criterion, and since the use of six rather than seven segments had important design advantages, the six-segment breakdown of the task and demonstration was accepted.

The practice and test materials consisted of ruler and compass and an Ozolithed booklet, the first page of which used a figure identical to that in the upper half of

[2] The film used was adapted from one developed for research use in a previous contract between the Air Force and Boston University, AF33(038)-22944 (A222).

Figure 1. Successive pages, the use of which will be described later, had figures with the essential parts of successive stages of the construction task added in.

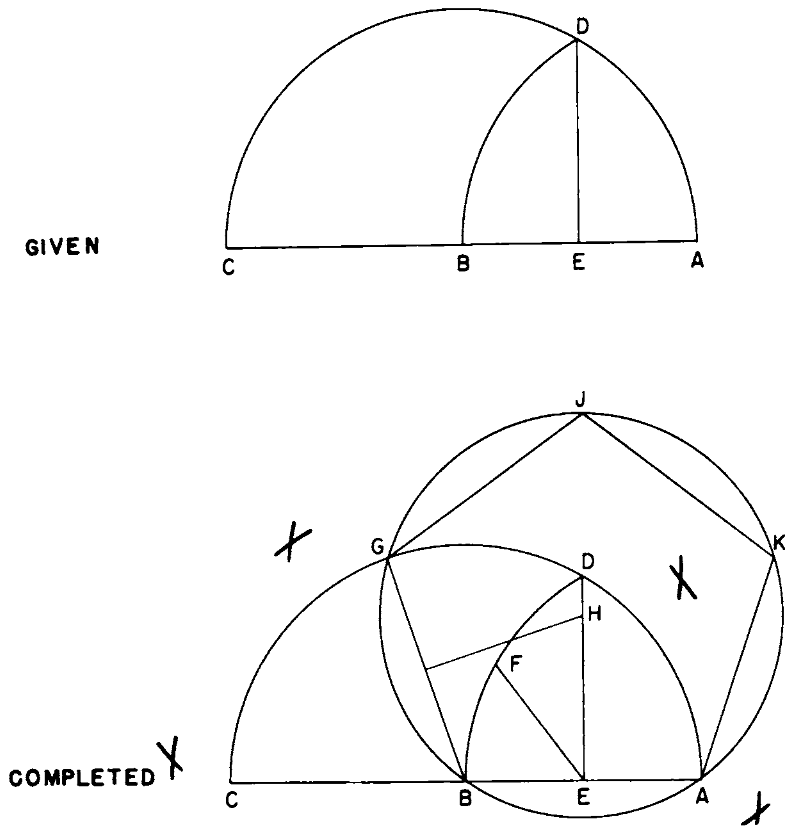


Figure 1. Construction Task

Subjects

The subjects were approximately 900 juniors and seniors in four high schools in the Boston area. Information was obtained from each subject concerning his background in geometry.

Design

All subjects viewed the entire film two times and were allowed to practice the geometric construction in conjunction with each demonstration.[3] Four treatments were utilized, as indicated above. These groups will be designated as follows:

[3]In the other experiment, using the mechanical-assembly task, three practices had been used prior to the final test. In the present experiment three practices were not feasible within a single high school class period.

Group I, Single D-A Span Segment group;
Group II-A, Two D-A Span group;
Group II-B, Three D-A Span group;
Group III, Whole Method group.

As indicated by these descriptions, Group I practiced after each D-A span segment; Group II-A practiced after the first two, second two, and last two D-A span segments; Group II-B practiced after the first three and last three D-A span segments; and Group III practiced only at the end of the entire demonstration.

The experiment was conducted during regular high school class periods with group administration of one of the four treatments in each of the classes involved. In all, 40 classes were used, these being randomly assigned to the four treatments in such a way as to provide 10 classes per treatment. In most of the statistical analysis the class was considered the sampling unit, giving a total N of 10 independent observations per experimental treatment.

Details of Training Procedure

D-A Span Segment Group. Each subject was equipped with a compass, ruler, and the practice and test booklet. An equilateral five-sided polygon was displayed on the blackboard of the classroom, and the class was told it would see a film demonstrating step-by-step how to construct such a figure. They were told that the film would be stopped after a part of the demonstration, at which time they were to open the booklet and attempt that part, working quickly but accurately and stopping work to attend to the next part of the demonstration as soon as the film started up again. They were informed in advance that the entire process would be repeated twice and that at the end they would attempt the entire construction without the aid of the film.

The film lasted four minutes and 30 seconds, and the six practice periods occupied a total of five minutes. The distribution of times for the six D-A Span segments and their corresponding practice periods are shown below.

D-A Span Segment	Film Time	Practice Time
1	53 secs.	55 secs.
2	12 secs.	15 secs.
3	46 secs.	55 secs.
4	58 secs.	75 secs.
5	25 secs.	40 secs.
6	75 secs.	60 secs.

The practice times used above had been determined in pre-testing a sample for performing the correct steps in each case. At the end of the time allotted for each practice period the subjects were told to turn that page of the booklet and watch the next part of the film. Each succeeding page contained the essential parts of the construction up to that point, so the subjects needed only to add in the next segment (the part just viewed) in each practice period.

At the end of the two demonstration-and-practice sessions the subjects were told to turn the booklet page again and attempt the entire construction. The booklet page for this complete attempt was exactly like that of the first practice page, and four minutes were allowed to attempt the entire construction.

At the end of this four-minute complete attempt, the subjects were asked to turn the page and start again from the point indicated, this being the same starting point as in the second page of practice (i.e., starting with the first segment already supplied). In the same way the subjects progressed a page at a time through all of the steps exactly as during the practice, and adhered to the same time schedule as during practice. The first part of this test will be referred to as the "free" test and the second part as the "guided" test.

Two D-A Span Group. The procedures of demonstration and practice involved the same total times as in the preceding group except that adjacent D-A span segments of the film were run in a continuous showing and practice was over two such segments at a time, giving the following distribution:

D-A Span Segments Combined	Film Time	Practice Time
1,2	65 secs.	70 secs.
3,4	104 secs.	130 secs.
5,6	100 secs.	100 secs.

The method of practice used with this group involved two stages, a first stage in which the subject attempted to perform the entire unit of the demonstration and a second stage in which he attempted only the second half of the unit. Thus in the very first practice the subject started with the same first page as in the Single D-A Span group already described. On this page he attempted everything shown thus far, namely the first Two D-A Span segments. He was then told to turn the second page of the booklet and finish the part of the construction indicated. This page was identical with the second page (i.e., second practice page) of the first group and was printed with segment #1 of the construction already completed. In a similar manner the Two D-A Span group's second practice began on page 3 of the booklet with an attempt to do both D-A span segments 3 and 4, but was followed by turning to page 4 of the booklet to attempt the fourth segment with everything up to that point completed for them in advance.

The purpose of this method of using practice time was to obtain not only a measure of how well the subjects could perform the unit of demonstration shown but also to measure their ability to perform each D-A span segment when testing conditions were the same as in the first group. Thus the subjects of the Two D-A Span group were tested on the second D-A span segment not only as part of their response to page 1 but also directly in response to page 2. The relative amounts of time permitted on these two successive aspects of each unit are indicated in the parenthesis of the foregoing breakdown of the distribution of time.

The final test procedure for the Two D-A Span group was identical with that for the Single D-A Span group.

Three D-A Span Group. This group was treated in a manner exactly analogous to the Two D-A Span group except for the use of three successive D-A span segments as the unit of demonstration and practice and the corresponding redistribution of times. An outline of the time distribution follows:

D-A Span Segments Combined	Film Time	Practice Time
1,2,3	111 secs.	125 secs. (60,20,45)
4,5,6	158 secs.	175 secs. (90,40,45)

The values inside the parentheses refer to the number of seconds allowed for page 1, page 2, and page 3 of the test booklet in the first unit, and for pages 4, 5, and 6 in the second unit. Thus, on page 1 the subjects had an opportunity to attempt all of the steps in segments 1, 2, and 3; on page 2 they had an opportunity to attempt the steps in segments 2 and 3; and on page 3 they had an opportunity to reproduce the D-A span segment 3 material. A similar arrangement held for the practice over segments 4, 5, and 6.

Whole-Method Group. The entire demonstration was shown to this group prior to each practice attempt. In each of its two practice periods the first 120 seconds was devoted to an attempt to reproduce the entire task depicted in the film. At the end of this period the remaining three minutes were subdivided as follows: 20 seconds on page 2, 35 seconds on page 3, 55 seconds on page 4, 25 seconds on page 5, and 45 seconds on page 6. The final test procedure was again identical with that used in the case of all of the other groups.

Results

Verification of D-A Span Segments

It will be recalled that the division of the learning task into six D-A span segments was based on a pre-test using Boston University students. The applicability of the same subdivision in the present high school sample can be seen in Table 1, which gives the percentages responding perfectly in the first practice of the indicated units of demonstration material. For Group I these units were single D-A span segments; for Group II-A they were double D-A span segments; for Group II-B they were triple D-A span segments; and for Group III there was only one unit, embracing the entire demonstration of six D-A span segments.

TABLE 1

Per Cent of Learners Able to Reproduce Correctly, Immediately After Demonstration, The Demonstration-Practice Units Used in Training

Demonstration Segment	Per Cent Perfect Per Demonstration Unit			
	Group I	Group II-A	Group II-B	Group III
1	66.8			
		44.7		
2	97.0		13.5	
3	85.5			0.4
		36.3		
4	77.9			
5	90.2		19.1	
		35.8		
6	90.6			

It can be seen in the table that, except for the first D-A span segment, the results conform to the plan of the study, which assumed that demonstration segments designated as a single D-A span in length would be perfectly reproducible, immediately after viewing, by more than 75 per cent but less than 100 per cent of the learners.

It can also be seen in Table 1 that the units used in Groups II-A, II-B, and III were too long to be reproducible by 75 per cent of the subjects; whereas each of the D-A span segments used in Group I was reproducible by over 75 per cent of the sample, except in the case of the first segment, which fell short by 8.2 per cent. In the pre-test sample this segment had been reproduced correctly by 88 per cent. The discrepancy appears to be due at least in part to the presence in the present sample of students having no background in geometry. The pre-test subjects were college students, all of whom presumably had taken the college preparatory course in high school. The high school sample, on the other hand, included both kinds of students—those taking the college preparatory course (which requires geometry), and those taking a course which did not include geometry. The results shown for the D-A span segment group (Group I), when broken down by the geometry versus no-geometry background, are as follows:

Demonstration Segment	Per Cent Perfect, Group I	
	Geometry	No Geometry
1	77.2	43.8
2	96.3	98.6
3	86.4	83.6
4	83.3	65.8
5	92.6	84.9
6	94.4	82.2

It is apparent in this breakdown that for the no-geometry subjects the first segment did not meet this study's definition of a D-A span segment. The same is true, but to a lesser extent, of the fourth segment. In each of these two segments the learner was required to use the compass in the operation of bisection, segment #1 being the bisection of an arc and segment #4 involving the bisection of a line. Partly in view of this non-homogeneity of the sample, the major analyses to follow were carried out separately for the geometry and no-geometry groups. It should be noted, however, that the observed differences between these two classes of subjects should not be attributed entirely to presence or absence of previous training in geometry, particularly since this appears to distinguish those in the college preparatory course versus those not in such a course in the high schools used. Thus the two kinds of students were even in separate classes for the most part in the present sample, all but two of the 40 classes used being either about 95 to 100 per cent geometry or 95 to 100 per cent no-geometry.

Training Effects

The chief measure used to determine the training effects of the various procedures during performance and practice was the average per cent, over the classes involved, getting the entire construction correctly completed. This measure was selected when it was discovered that the distribution of scores on the final test was grossly bi-modal, the two modes coming (a) at zero steps correct and (b) at all nine steps correct.[4] The extent of the bi-modality was such that 30 per cent of the subjects could not get past the first step and 43 per cent of the subjects could perform the entire task perfectly in the test. Thus only 27 per cent had graduated scores between these extremes. Thus, mean steps correct would not be a very meaningful measure of "central" tendency. As an alternative to mean number of correct steps, therefore, it was decided to use the per cent getting all nine steps correct as the measure of learning. Such percentages were determined for each of the separate classroom groups involved in the experiment, and mean per cents were obtained for each group, N being the number of classes upon which each average per cent was based.

Results for these average per cents for the first practice, second practice, and final free test are shown in Tables 2 and 3. Since the level of performance is quite different for the geometry and no-geometry subjects, the data are presented separately for these two sub-groups. The data in Tables 2 and 3 are derived solely from the first page of the final test in all groups and from the first page of each practice session in Groups II-A, II-B, and III. That is, the guided test results are not used in these tables, and the practice performance excludes that corresponding part of the practice in which the subject, after attempting on page 1 the demonstration

[4]It will be recalled that the task had nine different steps or operations, even though having only six D-A span segments.

unit just shown, tried the remainder of the unit from progressively more advanced starting points.[5]

TABLE 2

Average Per Cent of Geometry Subjects Completing Pentagon Perfectly

Group	Practice 1		Practice 2		Test	
1 D-A Segment	56.8	(7)	76.8	(7)	47.4	(7)
2 D-A Segment	19.4	(7)	52.3	(7)	67.6	(7)
3 D-A Segment	6.2	(6)	60.3	(6)	73.8	(6)
Whole	0.1	(7)	18.0	(7)	62.8	(7)

TABLE 3

Average Per Cent of No-Geometry Subjects Completing Pentagon Perfectly

Group	Practice 1		Practice 2		Test	
1 D-A Segment	25.9	(3)	55.1	(3)	9.8	(3)
2 D-A Segment	0.6	(4)	6.5	(4)	1.5	(4)
3 D-A Segment	0.0	(4)	2.2	(4)	0.0	(4)
Whole	0.0	(4)	0.4	(4)	9.0	(4)

Analysis for Geometry Subjects

An analysis of variance of the class per cent scores, using the arc-sine transformation, yielded significant treatment effects for both the first and second practices and an F just short of the five per cent level for the free test. Individual *t*-tests on the free test showed that the D-A Span group was significantly inferior to the Two D-A Span group and the Three D-A Span group, although not quite significantly inferior to the Whole group.

In this connection it should be noted that this inferiority of the Single D-A Span group comes about through loss of performance level when conditions are shifted from those of practice to those of testing. Thus the average per cent correct on all nine steps during the second practice is 76.8 for this group, a figure which drops to 47.4 under free test conditions. Using each class as its own control this drop of 76.8 - 47.4 = 29.4 has a *t* of 6.40, with $p < .001$ (both tails). By contrast,

[5]Results for the guided-practice data and guided-test data show substantially the same picture as those for the unguided test and practice data presented here.

the Double and Triple D-A Span groups show increases ($t = 2.05$, $P < .10$, and $t = 2.63$, $P < .05$, respectively) in the same shift from second practice to free test. All of the relationships described above can be seen in graphic presentation in Figure 2, which also brings out the fact that the Whole group started with essentially zero performance on the first practice, but showed striking improvement, particularly as a result of the second practice.

Analysis for No-Geometry Subjects

Turning attention from the "geometry" group to the "no-geometry" group it can be seen in Table 3 that there are some important similarities to the results just seen, but also some important differences. The major similarities are (a) the superiority of the Single D-A Span group during practice and (b) the huge drop in the performance of this group when conditions are shifted to the free test. Both effects are significant at the <.05 level or beyond. The major differences with the results for geometry subjects are concerned with the essential ineffectiveness of any of the procedures using more than a single D-A span as the unit of demonstration and practice. This can be seen graphically in Figure 3 in which the double and triple D-A span units give essentially zero performance in the final free test instead of crossing the Single D-A Span curve to a significantly higher value as in Figure 2.

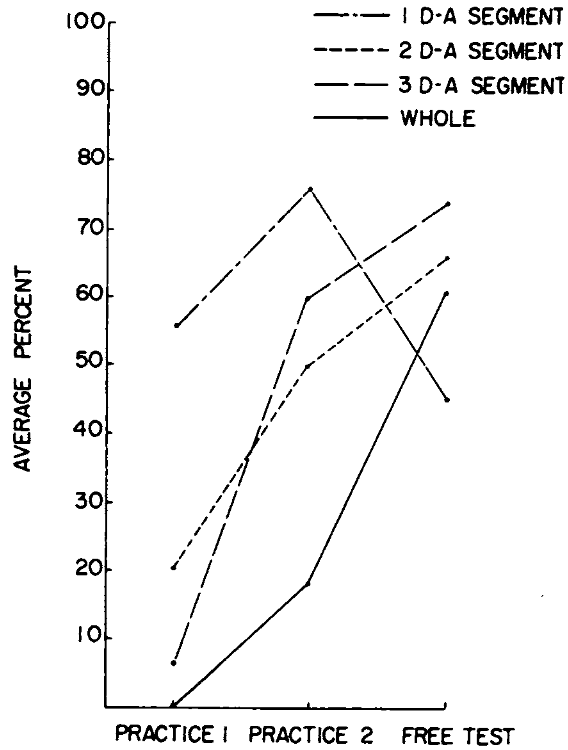


Figure 2. Average Percent of "Geometry" Subjects Completing all Steps Perfectly

Experiment 2

Purpose

A major purpose of Experiment 2 was to introduce two additional training procedures for comparison with those considered in the first experiment. One of these additional training procedures was the Transition condition in which the learner starts with smaller demonstration-practice segments and in successive showings increased to the Whole method.

The other was the Self-Pacing condition, in which the learner sets his own timing of demonstration-practice segments, stopping the demonstration to practice whenever he feels he has seen about as much as he can handle in practice. These

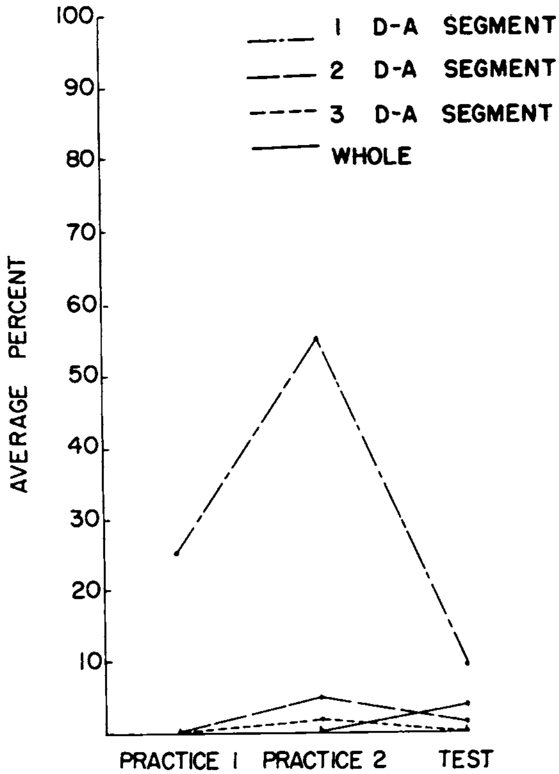


Figure 3. Average Percent of No-Geometry Subjects Completing all Steps Perfectly

correspond to Treatments 4 and 5, as identified both in the Introduction section of the present chapter and in the previous experiment with the mechanical-assembly task (Chapter 3).

A second purpose of Experiment 2 was to attempt to clarify the differences in results between the findings of Experiment 1 and those of the parallel study using a mechanical-assembly task (see Chapter 3). In the other study it had been found that the use of D-A span segments as the demonstration-practice unit gave superior performance, both in practice and in final test, to the use of larger segments, including the Whole method of training. In Experiment 1 of the present study, on the other hand, the use of D-A span segments gave superior performance only during practice with geometry subjects, and there was a large drop in performance under this method for both geometry and no-geometry subjects when conditions were shifted from those of practice to those of final test.

One possible interpretation of the relative success of the Whole method and the Larger-Segment methods in Experiment 1 of the present chapter is a procedural one. This interpretation centers around the possible practice advantage for these methods due to the procedure for getting a performance measure exactly comparable with that of the D-A Span Segment group. Thus the latter group of subjects always practiced on a page showing all the preceding steps at each practice stop. This gave them a possible advantage in cues relative to the Larger-Segment procedures, in which practice sessions had fewer cues for some of the steps, and relative to the Whole procedure, in which each practice session began with Page 1 of the booklet, providing a minimum of cues for successive steps. It will be recalled that, for purposes of getting comparable measures during practice, the initial attempts to perform what had just been seen in the Larger-Segment and Whole methods was immediately followed by going through successive pages giving the preceding steps exactly as in the D-A Span Segment group.

This procedure, introduced for the methodological purpose of getting comparable measures during practice, may also have had training effects which can be referred to as "guided practice" and which may have given a training advantage to the Larger-Segment and Whole method. The possible advantage is perhaps most clearly seen in the case of the Whole method. Here on each practice the subject was first given enough time to complete the entire geometric construction provided he knew the steps; then he turned the page and found the first step completed and was given at least enough time to go on to the next, and so forth. In other words, in addition to his practice attempts on the task as a whole, he had all of the practice steps of the D-A Span Segment group. Moreover, these latter steps come immediately after a practice attempt, so that to a certain extent the learner was getting knowledge of results and immediate correction. Partly to eliminate this factor the present experiment avoided this form of "guided practice" following attempts at larger amounts of the task. Instead the learner practiced only on a single page covering all of the demonstration unit just shown.

Another possible interpretation of the difference in results between Experiment 1 of the present chapter and the results of the study with the mechanical-assembly task derives from the fact that in the mechanical-assembly task there was greater correspondence between D-A span segments and "natural units" of the task. Prior to the pre-tests the assembly task was judged to have three natural units, and after the pretests it was decided that the first and last of these came close enough to the meaning of a D-A span that the division should be made at these points. The middle natural unit proved too long to be within the meaning of a D-A span segment, however, and it was subdivided into two D-A span segments; but the correspondence between D-A span segments and natural units was to the above extent fairly good. In Experiment 1 of the present chapter, however, the subdivision into D-A span segments was made on fairly empirical grounds, the definition being that at least 75 per cent of the pre-test sample should be able to perform perfectly after one showing of the segment.

Thus natural units were not considered in this subdivision, and it may be that the Two D-A Span segment and/or the Three D-A Span Segment groups in Experiment 1 of the present study involved subdivisions more closely corresponding to natural units. This could theoretically account for the general superiority on the test of these two groups.

In order to assess this possible factor in the discrepancy between the geometric construction results and the mechanical-assembly results, Experiment 2 utilized a group with subdivisions more along the lines of natural units independently of D-A span segments. The exact specification of natural units of the geometric construction task was more difficult than in the case of mechanical assembly, chiefly because there is no particular change of context from beginning to end of the construction. The decision made was that it could be subdivided into four natural units as follows (see Figure 1):

1. How to get distance FE from bisection of arc BD
2. How to get line GB from distance FE
3. How to get center of circle from bisection of line GB
4. How to mark off circle into polygon

There was some question as to whether parts 2 and 3 above actually should be separated, since to a certain extent they comprise a single unit of operations concerned with line GB. This question was resolved in favor of the above four units, however, partly because of the obvious parallel to the mechanical-assembly task, in which four segments were also used, although the middle two were more clearly a single natural unit in that task. The parallel also permitted running the present Transition group in the same manner as the Transition group of the mechanical-assembly task, i.e., first practice after each segment, second practice after each pair of segments, and third practice after the entire demonstration.

A final purpose in line with the same attempt to parallel the mechanical-assembly conditions was to utilize individual running and testing. The conditions of Experiment 1 were partly determined by the need to contain the training and testing within a regular high school class session. This had the disadvantage that training was limited to only two complete demonstration-practice sequences and that time scores could not be kept on individuals. In Experiment 2 all subjects were run on an individual basis, both time and errors were recorded, and it was possible to compute "performance rate" as defined in the mechanical-assembly study.

Method

Subjects

The subjects for Experiment 2 came chiefly from two sources—summer session students of Boston University Junior College, and students enrolled in an experimental psychology course in the Boston University College of Liberal Arts. These two kinds of subjects were run at different times (mid-summer and fall, respectively) and by different experimenters and will be treated in part as separate replications. The only additional source was a few graduate students included in the fall sample. In all, 65 subjects were tested individually, of whom 28 were Junior College students, and 34 were College of Liberal Arts students, and three were graduate students.

Design

The design is implied to a large extent under "Purpose" above, which notes that there was a Transition condition, a Self-Pacing condition, and a Natural Units condition. In addition, as a comparison group, a fourth condition was added which employed the Whole method like that used in Experiment 1 (except for an increase in number of complete practices and the elimination of the "guided-practice" feature of Experiment 1). All groups received three complete showings and three complete practices, the total practice time being the same (five minutes) for each group. The basic features of the design are given below, in which D stands for Demonstration, P stands for Practice, and 1, 2, 3, and 4 stand for the successive natural-unit segments of the task.

The film showing times for the four natural units were 65 seconds, 46 seconds, 58 seconds, and 100 seconds, respectively for parts 1, 2, 3, and 4. The corresponding practice times allotted to the natural units group (VI) were 70, 55, 75, and 100 seconds, respectively. The practice times for the Transition group (IV) corresponded to these same times in the combinations indicated in Table 4, i.e., P 12 permitted as practice time the sum of P1, P2, P3, and P4, i.e., 300 seconds. The Self-Pacing group offered special problems in timing since the subjects stopped to practice at unpredictable times. A prorated apportionment of practice time was worked out, based on the times permitted the parts in Group VI, to apply to the particular demonstration segments selected by a subject.[6]

TABLE 4

Design of Experiment 2

Group	Method	Trial	Procedure
III	Whole	1	D1, 2, 3, 4; P1, 2, 3, 4
		2	D1, 2, 3, 4; P1, 2, 3, 4
		3	D1, 2, 3, 4; P1, 2, 3, 4
		4	Free Test
IV	Transition	1	D1, P1; D2, P2; D3, P3; D4, P4
		2	D12, P12; D34, P34
		3	D1, 2, 3, 4; P1, 2, 3, 4
		4	Free Test
V	Self Pacing	1	D and P ad libitum
		2	D and P ad libitum
		3	D and P ad libitum
		4	Free Test
VI	Natural Units	1	D1, P1; D2, P2; D3, P3; D4, P4
		2	D1, P1; D2, P2; D3, P3; D4, P4
		3	D1, P1; D2, P2; D3, P3; D4, P4
		4	Free Test

The instructions to each subject in each of the groups were a straightforward explanation of what would be shown, how he would practice, and what he would be tested for at the end. In the case of the Self-Pacing group some decision was necessary about how much advice on pacing should be given. Each subject was encouraged to take as much as he could manage, but was cautioned against taking on

[6]For example, if a subject stops the demonstration at the completion of point G, his maximum practice time is 107 seconds. This is obtained by adding the 70 seconds for the first D-A unit to two-thirds of the second unit. Comparable proratings were applied to other stopping points.

so much that he would be at a loss when asked to practice. Otherwise the subject was told he was to be the judge of the length of demonstration. He was also told he could use the same or different stopping points on the second and third times through the demonstration.

One change in instructions from the Summer to Fall subjects was the addition of the statement that the subject should work as quickly and as accurately as possible. With both sets of subjects, however, a stop watch was obviously being used by the experimenter, and it is therefore doubtful if the added comment on working quickly was important.

Results

Results with the Junior College Sample

Since the results have a bearing both on questions raised about Experiment 1 and questions raised about the difference between the Experiment 1 results and the mechanical-assembly results, they are presented in terms both of per cent getting all steps perfectly (Figure 4) and of performance rate (Figure 5).

The results in Figure 4 should be viewed with caution because of the instability of percentages based on N = 7. They are presented in order to show the outcome in the same metric as in Experiment 1. It can be noted that, as in Experiment 1, the Whole method is the lowest during training. It is also lower under the present conditions than with the procedures of Experiment 1 (which utilized "guided practice"). This low level attains a certain degree of importance when one notes that in the third practice (after three demonstrations and two practices) the present Whole group achieves a perfect-on-all-steps level of only about 29 per cent as compared with the Experiment 1 figure of 62.8 per cent. A difference this large in this direction would occur by chance only about seven times in 100, which is not very reliable but is impressive considering that the present Whole group had an extra demonstration before their third free attempt. Thus the evidence tends to indicate some effect of the "guided practice" in Experiment 1.

Figure 5 shows the results in terms of the same metric used in the mechanical-assembly experiments, namely performance rate. This joint measure, utilizing number correct per unit of time, reflects improvement not only in the increase in number of correct steps but also in the reduction in time required to complete steps. The rate-score was obtained by multiplying number of correct steps by 60/time-in-seconds, which gives a score of number of steps correctly completed per minute during practice and final free test. This measure is much more sensitive to variations in performance than the dichotomous measure of "perfect" versus "not perfect," and means based on this measure are far more stable, even with only N = 7 at each point in the curves, than the per-cent-perfect values in Figure 4. The curves in Figure 5 are correspondingly more regular. Otherwise, however, the two figures present much the same pattern of results.

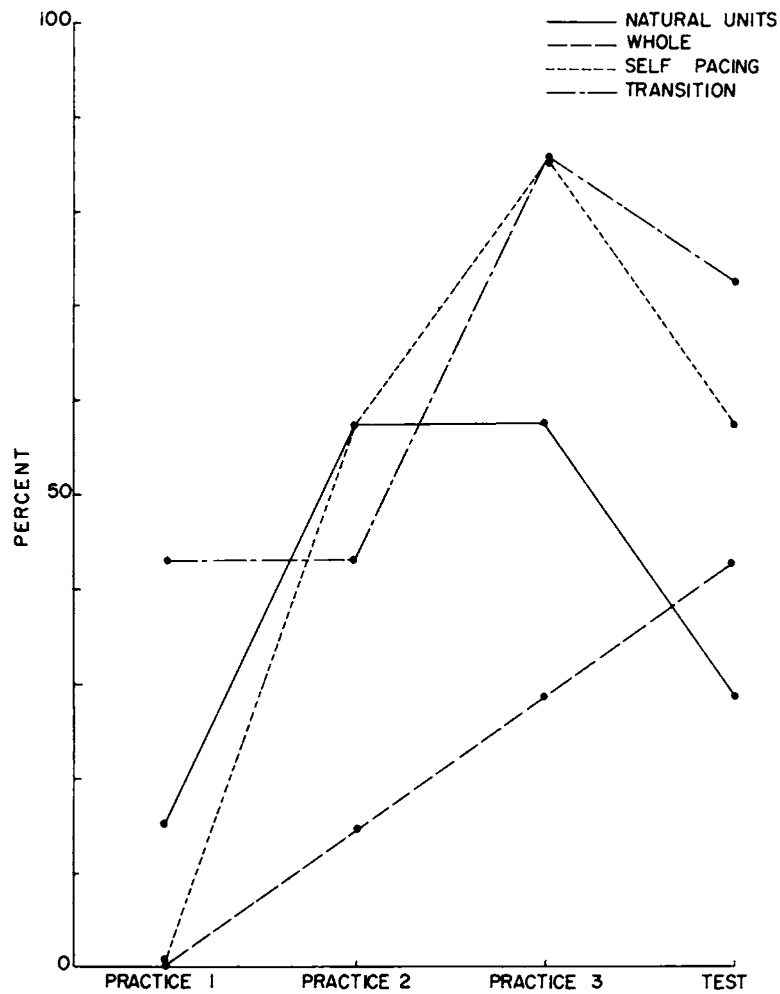


Figure 4. Per cent of Learners Getting All Steps Perfectly in Practice and Final Free Test. Junior College Students Only

If one sums the performance rates over the three practice trials for each subject, the F for treatment is 4.00 (3.01 needed at the five per cent level). None of the methods of interspersed practice differ significantly from each other, but the Whole method is significantly inferior to each of the interspersed-practice methods (Transition, Self-Pacing, and Natural Units). F for final test comparisons among the four groups is highly significant, and $\underline{t}$ is significant at the .001 level for Transition versus Natural Units and at the .01 level for Transition versus Whole method.

One of the most striking features of Figure 5 is the sharp drop in the Natural-Units group in the shift from third practice to final test. This drop is highly significant, and corresponds in pattern to the drop found for the D-A Span Segment

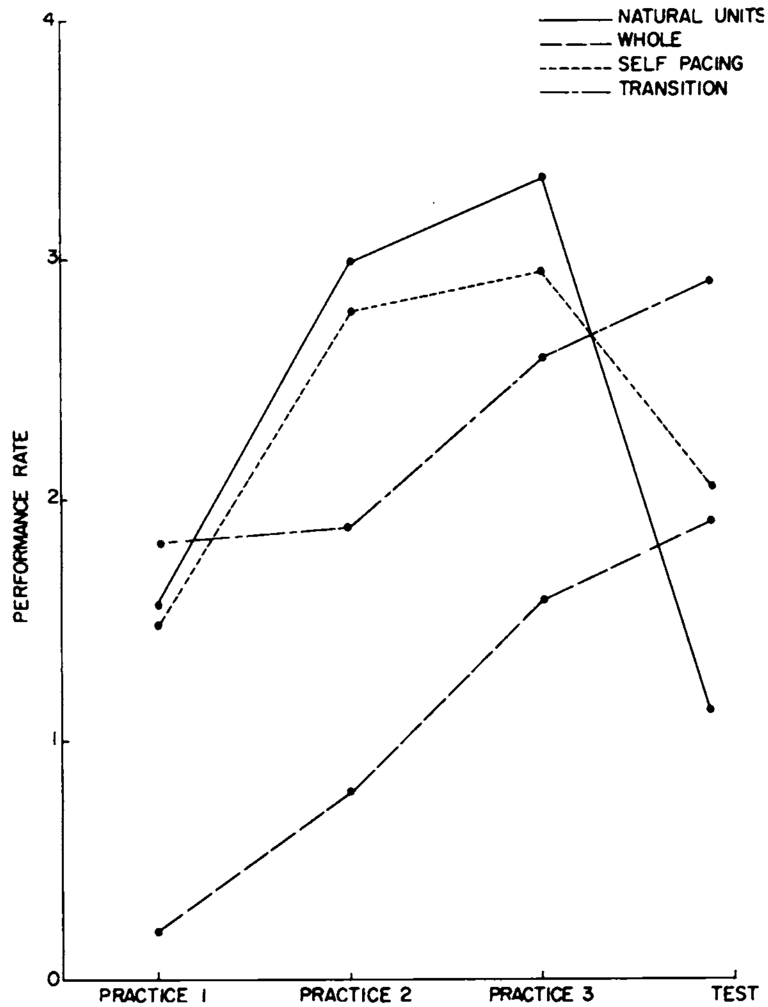


Figure 5. Performance Rate (Steps Correctly Completed Per Minute) in Practice and Final Free Test. Junior College Students Only

group in Experiment 1. It indicates that the natural units used in Experiment 2 did not circumvent the comparable difficulty found in the high school sample with empirical D-A span segments.

Results with the More Advanced Sample

Results with the Liberal Arts College and Graduate School sample were not very comparable to the foregoing findings with the Junior College sample. Figures 6 and 7 show the per cent getting all steps perfect and the performance rates, respectively. The only significant effect in Figure 7 is the lower level during practice for the Whole group compared with the others. One possible conclusion appears to

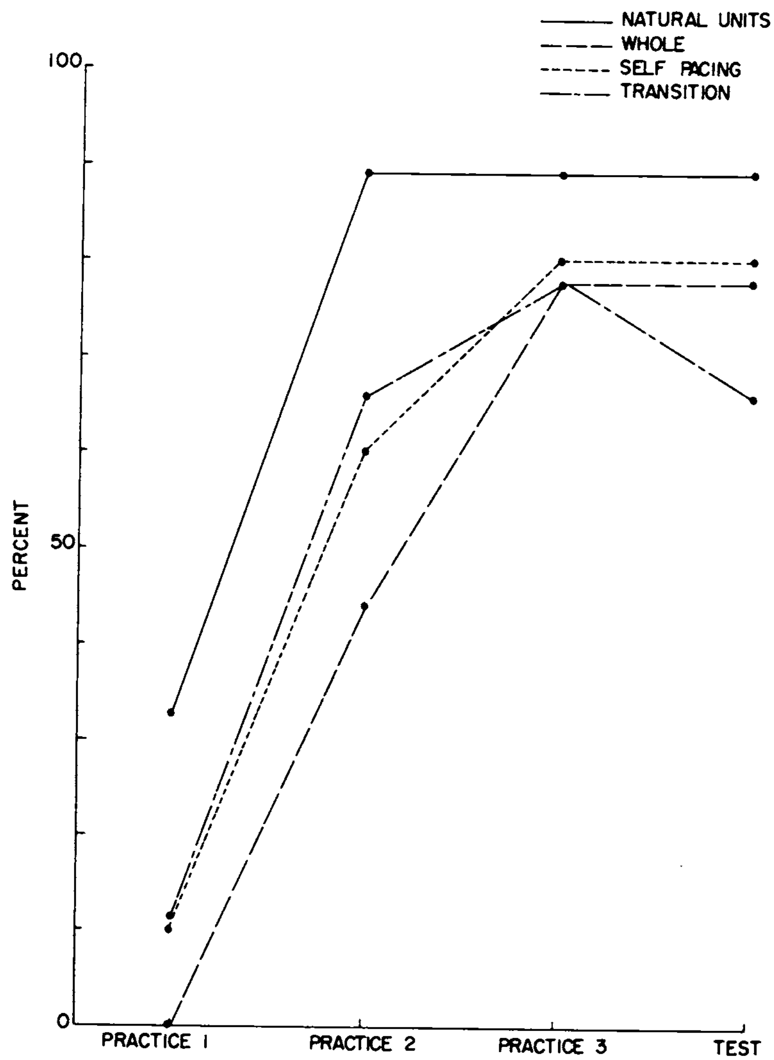


Figure 6. Per Cent of Learners Getting All Steps Perfectly in Practice and Final Free Test. Liberal Arts and Graduate Students

be that these subjects were such good learners that nearly all of them were able to attain high scores with the amount of training given regardless of method of demonstration and practice.

Interpretation

The results of Experiments 1 and 2 are contrary to any simple expectation of improvement in acquisition and/or retention of demonstration content as a result of introducing overt practice whenever a D-A span has been reached. The most striking

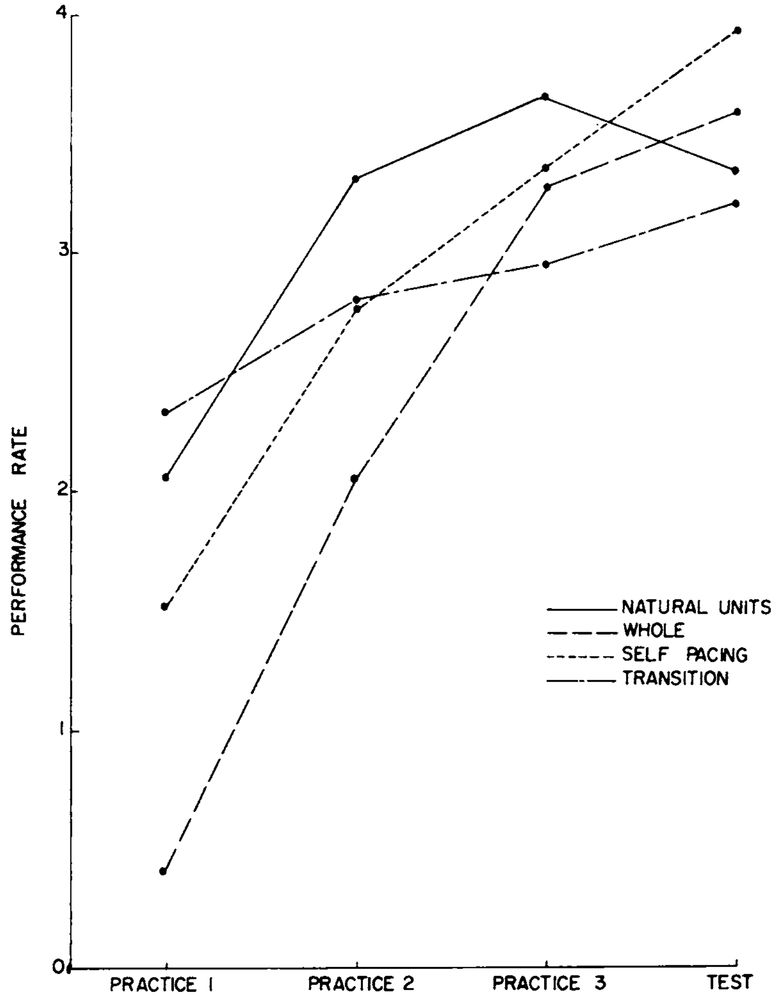


Figure 7. Performance Rate (Steps Correctly Completed Per Minute) in Practice and Final Free Test. Liberal Arts and Graduate Students

feature of the results is the tendency for smaller-segment forms of practice to give good performance during training but to show a deterioration in performance in the final test, in which there is no immediate support from the demonstration. This was found in Experiment 1, where the smallest segments were empirical D-A span segments, and it was found—at least for the Junior College sample—where the smallest units were "natural units" of the task. Even in the more advanced sample, made up of Liberal Arts College and Graduate School subjects, the Natural Units treatment was the only one which declined rather than improving from last practice to test performance. While this latter differential result was not statistically significant in this sample, it is in line with the rest of the findings. Since part of the purpose of Experiment 2 was to check on possible procedural explanations of the corresponding performance drop in the D-A Span group of Experiment 1, and, since Experiment 2 in general showed a similar outcome, the conclusion appears warranted that for this geometry task some form of Larger-Segment practice is desirable for transfer to the test conditions of actual performance without support from the demonstration.

The Transition procedure is theoretically the most advantageous admixture, but while it turned out superior with the Junior College students in Experiment 2, the superiority was not very reliable. There seems to be a good possibility, however, that the transition in this experiment went at either the wrong rate or in the wrong units. When the learners were shifted from single to double "natural units," their performance failed to increase with practice in terms of per cent perfect and actually dropped in mean number of steps correct, the rise in performance rate coming mainly from an increase in speed. The units combined in the transition may not have been the most judicious from the standpoint of those parts of the task that naturally would become integrated early. The 4-2-1 transition used had definite advantages from the standpoint of the experimental procedure and was not strongly dictated by the content of the task.

It should be noted that the Self-Pacing procedure, as handled by the subjects, was itself a Transition procedure. Thus the subjects uniformly chose to start with smaller units and progress to larger ones. Transition comes naturally to the learner because he needs less prompting from the film as his performance improves. It is interesting that the Self-Pacing procedure was less effective with the Junior College students than with the Liberal Arts and Graduate students. This suggests that the presumably more intelligent and able may be better judges of how to pace themselves and combine elements that integrate naturally. It was also interesting that superior students tended to use a 3-2-1 sequence rather than the 4-2-1 imposed on the Transition group. This tends to confirm the early suggestion that the task may have three rather than four natural units, and that the combinations used in the present Transition procedure actually split up a middle natural unit.

It appears likely from the present study that the method of practicing exclusively in single D-A span segments, where these are not themselves in need of internal integration, will have the effect of detracting from the integration of the task as a whole. The task is never practiced in actual sequence because of the interruption of the demonstration in order to practice and the interruption of the practice in order to watch the next phase of the demonstration. The optimum procedure appears to require some admixture of the advantages of consecutive performance and the advantages of being able to recall what is to be performed. In Experiment 1 both the Single D-A Segment and the Whole method appeared to be inferior extremes, the former giving a sizable decrement when the continuous performance of the free test was required and the latter producing a steady rise but not to a very high level.

The results with the geometry task are quite different from those with the mechanical-assembly task (Chapter 3), in which the Smaller-Segment method was superior to both the Larger-Segment and the Whole methods of practice in the final test as well as during training. Since the geometry task results failed to show this even when, as in Experiment 2, an effort was made to make the procedures of practice and testing as comparable as possible for the two kinds of tasks, the conclusion appears to be that the discrepancy involves inherent differences in the two kinds of tasks.

One such difference is that the assembly task is a much longer sequence of steps, that is, the steps are more numerous. Thus the assembly task involved assembling—in order—thirty different parts, whereas the geometry task breaks down into only nine discrete steps. By "hindsight" it might be argued that the geometry task is not actually the kind of complex sequence to which the theory applies. That is, the theory applies where complexity derives from length of task measured in number of serial elements. The geometry task is complex, but its complexity does not derive from the large number of elements that must be strung together in sequence, as was intended by the "consolidation" concept discussed in the Introduction section of Chapter 3 (see also Chapter 2).

Another difference that may be of importance is that the geometry task requires the subject to perform without any external reminders of what he is supposed to do. It is a straight recall task as contrasted with the task of the automobile distributor, in which all of the parts are in evidence and subject's job is to use them in the proper order. It may be that for this reason the geometry task requires the subject to anticipate subsequent steps in order to execute the immediate step (e.g., he must recall that he is to bisect a line in order to remember that he should pick up the compass and draw arcs). This would stress the need for practice in larger segments which cover more than one or a few steps at a time.

The main conclusion from the present results, taken in conjunction with the prior results from the mechanical-assembly task, is that the optimum method of utilizing overt practice in connection with a demonstration is not necessarily the method which maximizes correct performance of the overt responses during practice. The D-A span unit, as the practice unit, uniformly gives superior performance during practice but is not guaranteed to connect together the different parts of the total task. A compromise which gives at least some practice at larger units of the task may be desirable even if it entails more errors in performance at the outset of practice.

CHAPTER 5

COMBINING PRACTICE WITH DEMONSTRATION IN TEACHING COMPLEX SEQUENCES: SUMMARY AND INTERPRETATION[1]

Nathan Maccoby
Fred D. Sheffield

The aim of the present chapter is to summarize the purposes, theory, procedures, and results of the studies reported in the two preceding chapters, and to provide interpretations of the results, considering both studies jointly.

The immediate objective of the research was to determine the effects of length and distribution of overt practice periods on learning serial manual tasks from repeated film showings demonstrating the task, the showings and practice to be combined in a single training session. While the setting for the research was thus the joint use of filmed demonstrations and overt practice, the theory and results are not considered specific to filmed demonstrations. Numerous other training situations involve a demonstration of correct performance, followed by an attempt on the part of the trainee to reproduce in overt performance what he has passively observed during the demonstration. Typically such demonstrations are "live," and "live" demonstrations may provide a more realistic setting than filmed demonstrations in which to apply any conclusions of the present research, since the usual conditions of training-film projection (e.g., assembling trainees in a theater or projection room) make the interspersing of overt practice periods difficult. When and if portable, automatic-loading, sound-film viewers come into general use, this difficulty will be minimized. From a research standpoint, however, the use of a filmed demonstration had the highly desirable feature that the demonstration was completely constant from one presentation to the next.

It should be noted that the research purpose was not concerned with assessing the gains from overt practice *per se*. Pure demonstration rarely if ever suffices to provide complete learning and is normally considered only part of the teaching process where ability to perform perfectly is the ultimate criterion of learning. In line with this general fact, the present research assumed that overt practice would add considerably to the training, however this practice was administered. Experimental comparisons to prove this point were not included in the research on the

[1]This chapter summarizes the conclusions and interpretations derived from work conducted under Air Force Contract AF 18(600)-1210 at Boston University through theoretical and experimental work reported in Chapters 1, 2, and 3. A preliminary summary of some of the experimental findings has been previously reported by Maccoby and Sheffield (1956, 1957, 1958).

grounds that the findings would not be very useful. The only possible value of using repeated showings of only the demonstration would be to establish a "demonstration-only" baseline against which to assess the gains from practice and the relative sizes of effects due to different methods of utilizing practice. Such a baseline was not used, however, because it was decided that the quantitative results that would emerge from such comparisons would perhaps be misleading with regard to the effects of successive viewings of a film without practice. The reason for this conclusion was that previous experience indicates that unless some strong external motivation can be maintained, learners get so bored with repeat showings of the same film that they no longer pay attention. The training effects of a single demonstration without any practice are assessable from the present studies.[2] The result that would have ensued from more than one showing without practice was not investigated for the reasons just mentioned. Thus the experimental comparisons always used both demonstration and practice, and the major variable was the timing of these two factors.

At the simplest level, the theoretical considerations involved in the research centered around the expectation that, in the case of a lengthy and/or complex serial task, attempted practice after a brief segment of the total demonstration will result in more accurate overt performance of that segment than attempted practice only at the end of the entire demonstration. That is, if practice came only at the end, the learner would forget a large part of what was shown by the end of a complete demonstration and would be unable to perform when called upon to practice. If, on the other hand, he were given a time-out for practice at a number of different stopping points in the demonstration, he would be able to reproduce almost perfectly what he had just seen demonstrated in the preceding section.

If accurate practice were the only consideration, the implication would be that overt practice should be interspersed throughout such demonstrations. Another consideration has to be taken into account, however; namely, that the learner has to integrate the total learning task into a complete serial performance which he can execute from beginning to end without being prompted by the demonstration. The research was predicated on the expectation that some optimum compromise could be found between these two considerations, one of which suggests a number of interspersed practices and the other practicing only at the end of the total demonstration.[3]

Two main kinds of potential stopping points for interspersed practice were considered, those involving a "demonstration-assimilation span" and those involving "natural units" of the task. The former kind of stopping point (D-A span) is one which comes at a point when the learner has assimilated about as much demonstration material as he can hold in his mind well enough to translate into overt practice with

[2] See performance at the end of the first showing in the Whole Method, Chapters 3 and 4.

[3] See Chapter 2 for a more complete discussion of theoretical factors.

little or no error. If the demonstration were subdivided in this way, the learner would be able to perform almost perfectly in his overt practice sessions because each one would cover a small enough portion of the demonstration that no forgetting (interference) would be encountered. The other kind of stopping point (natural unit) comes at the end of an organized sub-unit of the total task. It may be longer than a D-A span, so that the learner cannot practice at all perfectly after a single viewing of that sub-unit of the demonstration, but because of its potential for integration into a unified component of the total task, it will be less subject to interference from other parts of the task after two or more complete demonstration-and-practice sessions.

A word should be said about how a D-A span was defined when dealing with groups of individuals of varying ability. Since better learners have longer D-A spans it was necessary either to have different stopping points for different individuals or to have one set of stopping points that made segments somewhat more than a D-A span for some individuals and somewhat less than a D-A span for others. The latter course was chosen, partly because it is more realistically related to the way training films are ordinarily used, i.e., one set presentation before a group of individuals. In line with this the operating definition of a D-A span segment in the present research was a segment which could be practiced with a mean group-performance score of about 75 per cent correct responding immediately after viewing.

In the conduct of the research the smallest demonstration-and-practice segment used was the D-A span length. Presumably stopping for practice after smaller segments than this would involve a loss in the overall serial integration of performance without much gain from accuracy of practice. At the other extreme the largest demonstration-and-practice segment used was the entire task (i.e., demonstration all the way through before practice), which was called the "Whole method." The compromises used were of several kinds but basically involved one or another of the following: (a) demonstration-practice segments made up of two or more adjacent D-A span segments, (b) demonstration-practice segments made up by combining two adjacent natural units, and (c) transition, on successive showings, starting first with single D-A span or natural-unit segments and then advancing to combinations of these, and finally to the Whole method. Another compromise procedure, similar to (c) above, was the Self-Pacing condition, in which the learner himself decided when to stop the demonstration and intersperse an overt-practice period.

With the mechanical-assembly task used, natural units could be designated rather unambiguously in terms of the physical integrity of the sub-assemblies treated as natural units, i.e., the parts of each such unit were assembled together into a single substructure different from that of the other units. This subdivision produced three natural units, the first and last of which were approximately of D-A span length. The middle one of these, however, was more difficult than the first and last and was divided into two approximate D-A span segments, giving four demonstration-and-practice segments. Using these segments and three complete showings, it was found that interspersed practice was significantly superior to the Whole method in both time and error scores and in performance rate (number of correctly assembled parts per unit of time attempting performance). The same breakdown of sub-units was also superior to the use of larger sub-units formed by combining the first two and last two of the above four segments into two "halves" of the task, with stopping points for practice coming at the end of each "half." When compared with the Transition method, however, the four-segment method was not superior, final test scores being very similar for these two.

These results clearly demonstrate a case in which the interspersing of practice during the demonstration has a definite advantage over using the same amount of practice time at the end of each complete demonstration. The results showed the expected superiority of the Transition method in the final test, but for this task this superiority was not statistically significant in the comparison between the Transition method and the use of the four segments subdivided according to D-A span. This suggests the possibility that as a general rule the use of transition, which should produce gains in serial integration from advancing to progressively larger units, will commonly suffer a compensating loss of integration within natural units. There are extenuating circumstances in the present results, however, perhaps the most important of which is the fact that the transition sequence used in the experiment did not permit integration of the middle natural unit until the Whole method was used in the third showing. This latter interpretation is supported by the fact that the Transition group showed a rise from third practice to test, whereas the four-segment group showed a drop from third practice to test.

One point needs special emphasis in the present summary of the mechanical-assembly results; namely, the data do not provide evidence that the smaller the segment of demonstration and practice the better the learning. This generalization of the data would be superficially correct, but only because the smallest segments investigated were either natural units or D-A span units. The presumption is that if these had been further subdivided into more segments, the training outcome would have been inferior. Presumably, relatively little would be gained in accuracy of practice performance in such a case, and a considerable loss in the integration within natural units and in the serial integration of the total task would be expected.

Another point deserving mention is that it was partly an accident of the present task that the first and last D-A span segments were also natural units. Part of the superiority of the four-segment method may have been due to having two stopping points coincide with the end of natural units rather than being entirely due to the use of D-A span segments. Thus the dimension along which the present findings should be generalized is not entirely clear.[4]

The Self-Pacing data in the mechanical-assembly experiments suggest that no purely ad libitum Self-Pacing method will be very successful. Unguided self-pacing results in such a great variety of stopping points that these cannot be regarded

[4]This distinction was perhaps not made sufficiently clear in Chapter 4, which most often refers to the four sub-divisions as D-A span segments. They are approximate D-A span segments but three of the four stopping points also coincided with the end of natural units.

as accurately adjusted to the abilities of the learners. Rather, it appears that learners under these conditions adopt private theories as to the best method of distributing demonstration and practice; these private theories would not be expected to be appropriate in a very large proportion of cases.

The major possible advantage in a Self-Pacing procedure rests in the fact that individuals differ in their personal D-A spans. If the learner could accurately assess when he has "seen enough," he could stop the demonstration and shift to overt practice at points appropriate to his own ability level. This might be a superior learning procedure provided the learner were at the same time given certain rules to follow in his selection of stopping points. Thus he could be made aware of the natural units of the task, which are independent of individual D-A spans, and instructed to make a practice stop at the end of each natural unit. He also could be instructed to attempt larger segments of demonstration with successive complete showings in order to foster serial integration of the total task. Thus, some form of guided self-pacing might turn out to be the most effective training method if self-pacing is at all feasible.

It is of some relevance that in the geometric-construction task used in the present studies the Self-Pacing method was very effective when used by superior learners (Boston University Liberal Arts majors and graduate students, as contrasted with Boston University Junior College students). With the lower ability group there was a drop in performance from the last practice to the final test, indicating that these subjects had failed to provide sufficient transition to the total task in the pacing used. The superior learners, on the other hand, showed continuous improvement in performance rate, indicating adequate transition from training conditions to final test on the total task. Their pattern of distribution of demonstration and practice was in line with their performance; that is, most of them utilized a transition procedure which involved viewing and practicing the entire task without interruption on the last training session. Thus it may be that superior learners (or intelligent students) would use appropriate self-pacing techniques even with little or no guidance.

It should be pointed out that self-pacing is not always feasible even if it were to produce superior learning. A good deal of training is conducted in classes or other groups in which all individuals have to follow the same pacing of demonstration and practice. Also, in homogeneous groups, there would be no need for self-pacing because D-A spans would vary little from one individual to the next. In any case, self-pacing is not a genuine training procedure. It is a method of adjusting the distribution of demonstration and practice to the abilities of the learner and, as such, is not expected to be superior to methods of controlled pacing which are also adjusted for learner ability but which are selected by the trainer rather than the trainee. Similarly in situations in which self-pacing is inescapable, it would be expected that rules for guided self-pacing would produce better learning than *ad libitum* distribution of demonstration and practice.

The discussion thus far has centered primarily around the results of the experiments using the mechanical-assembly task. In this task there was more correspondence between natural units and D-A span units than in the case of the geometric-construction task. The results of the experiments with the geometric-construction task indicate, however, that generalization should probably not be made along the D-A span continuum. One of the most important results of the geometry-task experiments was that, whereas practice performance was maintained at a high level during training by D-A span segments, there was apparently a considerable sacrifice in serial integration of the total task. This is supported by large drops in performance level when the learners were shifted from training conditions to the conditions of final test on the task as a whole. However, in the group experiments using high school students with some previous geometry, the compromise methods involving segments of two and three D-A spans in length, while producing inferior performance during parts of training, produced superior performance on the final test.

These results suggest that some procedure requiring the learner to take on more demonstration than he can actually reproduce perfectly in practice will have a better overall effect. Presumably this is due to the gains in serial integration provided by the greater continuity of larger segments. At the same time, however, the complete continuity of the Whole method did not prove superior to the compromise methods using two and three D-A spans, presumably because the forgetting caused by intra-serial interference prevented learners from performing adequately during practice.

Another possible factor in the above results is that of integration of the material along lines of natural units. The breakdown of material into D-A span segments will not necessarily result in stopping points which coincide with natural units. Thus, stopping at the end of an empirically determined D-A span may impede the integration within natural units which would otherwise take place and facilitate learning. From this standpoint all of the Larger Segment methods (including the Whole method) have an advantage over the D-A Span Segment method, the reason being simply that with fewer stopping points there is less chance of breaking into a natural unit. The import of this line of reasoning would be that if the task is complex or lengthy, the Whole method will be inferior simply because the learner cannot retain enough of the demonstration to profit much from practice, but the D-A Span Segment method may be inferior because the learner cannot integrate natural units of the task. If the natural units of the task cannot be identified, therefore, the best procedure would be a compromise in which the demonstration-and-practice segments used are larger than single D-A span segments. If natural units can be identified, on the other hand, they would be the appropriate segments for initial practice.

The above generalizations can also be considered appropriate as applied to the geometry-task data from high school students with no background in geometry. These students also were, for the most part, not preparing for college in their high school courses, and therefore probably differ in other respects than the lack of a geometry course in their educational background. Whatever the reasons, these subjects performed at a much lower level than the "geometry" subjects, and for them the D-A Span Segment method comes out ahead of all of the other methods both in performance during training and in performance on the final test. Superficially this might be interpreted to mean that for learners of low initial ability (in terms either of prior familiarity with the material or low native talent), practice after single D-A span segments works best. For these students, however, two of the six D-A span segments used in the demonstration failed to meet the criterion of a D-A span. The six segments had been selected on the basis of the performance of a pretest sample of Boston University college students; and, while they turned out also to be appropriate for the geometry students in the high school sample, they fell short in this respect for the non-geometry students. Thus the D-A span segments for the latter group were to this extent already a compromise in which the learners were being pushed beyond their actual D-A spans in the stopping points used.

It should be emphasized that the gain in training effectiveness from any compromise method is theoretically brought about by the increased serial integration of the task. Such serial integration involves the connecting of subsequent responses in the sequence to the cues produced by antecedent responses in the sequence. The learner must know what to do next at every stage of performance, the cues for such knowledge coming primarily from what he has reproduced thus far. A difficulty with any segmental practice procedure is that the learner completes a segment and then, instead of doing the next thing, in the sequence, he turns to the demonstration to observe what is done next. Thus performance of a segment gets cued to the demonstration rather than to the preceding performance, and unless the learner at this point in attempted total task performance can symbolically or perceptually reproduce the content of the demonstration, he will be unable to move to the next segment.

It is for this reason that Transition methods were expected to prove superior to others. With transition, the learner progressively increases the size of segments, presumably moving from segmental performance to total performance by steps which maximize both the ability to reproduce everything in the demonstration and the serial integration of the overt performance. In the mechanical-assembly task and in the geometric-construction task with the Junior College sample, the Transition procedure was superior to natural units in the final test, but only in the latter case was this superiority large enough to rule out the possibility of a chance difference. Transition may be regarded as another form of compromise between the gains from accurate practice that come from the use of smaller segments and the gains from serial integration that come from the use of larger segments.

Some comment should be made concerning the attempt to use "natural units" in the case of the geometric-construction task. It was thought that by changing from the six D-A span segments used in the high school sample to the natural-unit segments used in the Junior College sample there would be a smaller drop in performance when conditions were shifted from those of training to those of testing. The expectation was not confirmed in that a large drop was still in evidence with the four segments that were identified as natural units. This may be partly due to a faulty choice of natural units, in that it is somewhat unclear as to whether there are four or three such units in this task. In this difference in possible choice of natural units, however, the boundaries unfortunately do not coincide at all. Thus if the three-unit alternative is the correct one, the four-unit selection used is wrong in all of its stopping points except the final one. The issue perhaps could be settled only by further research, but it brings out the point that, at the stage of development of the concept of "natural units" reached in the present research, these are not unambiguously identifiable.

It is relevant to point out in this connection that in the most intelligent and knowledgeable sample tested on the geometry task (the liberal arts students taking a course in experimental psychology plus a few graduate students in psychology), the self-pacing subjects tended to use a three-stop method rather than the four-stop version employed by the experimenters. This suggests not only that there actually were three rather than four natural units, but also that a possible method of empirically determining natural units of a task would be to observe the manner in which a small group of a priori good learners pace themselves as between demonstration and practice.

An alternative interpretation of the loss in shifting from the four natural units to the final test on the task as a whole is, of course, that there were no genuine natural units in this task, making the large loss simply a decline due to practicing in too-small segments, with consequent sacrifice of serial integration. The simplest index of a natural unit is the constancy of context in which a part of the task is practiced. In the mechanical-assembly task, for example, a series of parts which are all attached to the same basic piece make an obvious natural unit. The components make up a unit which has the constant cue during assembly provided by the basic piece to which they are attached, and when the sub-assembly is completed it can be seen, handled, and inspected as a genuine physical unit. In the geometry task there was no real counterpart of a sub-assembly, and a single fairly constant cue-pattern was present throughout the entire construction. Thus it may be that, whereas the construction of a five-sided equilateral polygon might itself be a natural unit in some larger geometric task, it does not readily break down into natural sub-units.

Another alternative interpretation of the discrepancy between the assembly results and the geometry-task results must be considered. This is the alternative that the two tasks involve inherent differences in the nature of the learning involved. Thus the facts are that in the final test neither the D-A span method with high school students nor the natural-units method with Junior College students came out superior to the Whole method with the geometry task. This is to be contrasted with the clear superiority of the D-A Span Segment and/or Natural Units method over the Whole method in the case of the mechanical-assembly task. This difference could be attributed to the differences in kind of task. Several such differences can be noted; the ones that suggest explanatory hypotheses for the difference in results are discussed below.

One difference is that in the one task (mechanical assembly), the parts of the task are all suggested by the physically present parts of the assembly; whereas in the other task (geometric construction), the parts of the task have to be suggested by the learner's associations. That is, in the one case the disassembled parts have to be rearranged into the correct assembly; whereas in the other case no parts are present; the learner must supply them on the basis of sheer recall. In the latter case, it may well be that in the geometry task the learner has more need for a perceptual impression of the final product from which he can select the appropriate components in the construction one at a time in the proper order. To the extent that this is the case, taking time out for overt practice is to some extent a distraction. The best training for the perceptual task may be uninterrupted observation of the demonstration itself, the chief value of practice being to bring out mistakes and show the learner what to look for next time. This interpretation suggests that, with one or two more practices than used in the geometric-construction study, the Whole method would have outstripped the others.

A related difference is that the final product in the geometry task is completely visible because of its two-dimensional nature, whereas most of the mechanical assembly is hidden when it is completed. A good student in geometry, knowing that the purpose of the task was to make an equilateral pentagon, could merely study the end product and probably perceive and memorize the steps needed. By contrast the mechanical assembly requires seeing the steps executed in the sequence, because each immediate step tends to hide preceding steps and future steps cannot be seen until they take place.

This consideration leads readily to another important difference, namely that the mechanical-assembly task is a serial-learning task of 30 steps, whereas the geometry task has only nine steps and, as noted above, has a somewhat dubious status as a serial-learning task in the usual sense. The experiments (and theory) are couched in terms of complex response sequences, the main theoretical factor making learning difficult being the expected intra-serial interference taking place with a lengthy sequence. The geometry task is admittedly complex, and must be done in a sequence. But it is not a lengthy sequence, and its complexity is not the complexity that arises from sheer length and similarity of elements in the sequence. This interpretation suggests that the theoretical analysis appropriate for lengthy mechanical assemblies and other lengthy serial tasks is not appropriate to the geometry task. Rather, the latter task is one which calls more for what in Chapter 2 has been called a "perceptual blueprint."

As a final consideration, it should be pointed out that the orientation of these studies has tended to be toward perfect performance as the final product of the training. Different sorts of conclusions might arise if the goal were the maximum preparation possible in a given period devoted to training. In the mechanical-assembly task, for example, the performance level is close to perfect at the end of training in all groups, but the amount of time devoted to training is considerable. If only a relatively small part of the training were to come from mixed demonstration-and-practice sessions, revisions in recommendations might be necessary. From the present results, the only thing that can be said is that in the course of a single showing the D-A Span Segment groups perform at a much higher level. They have seen the entire demonstration and practiced most of the overt performance without many errors. They are therefore probably better prepared for future training than groups that are trained according to methods that pay off in the long run but start with a low overt-performance ability. Said another way, the conclusions reached here might be quite different if only one showing had been utilized and all groups had been given the final test after this single showing.

CHAPTER 6

REPETITIVE VERSUS CONSECUTIVE DEMONSTRATION AND PRACTICE IN THE LEARNING OF A SERIAL MECHANICAL-ASSEMBLY TASK[1]

Garry J. Margolius
Fred D. Sheffield
Nathan Maccoby

Part 1: Experiment Using the Automobile Distributor-Assembly Task

Background and Purpose

One of the well-established conclusions from studies of human serial learning is that the longer the sequence to be learned, the greater the number of repetitions of the entire task required to reach a criterion of mastery. This relation does not hold for sequences within the "memory span," in which case a single repetition is sufficient, but beyond the memory span more repetitions per item in the sequence are required as the sequence is lengthened. This relation looms in importance when there are practical reasons (e.g., military significance) for mastering an inherently lengthy serial-learning task (e.g., assembly of an intricate mechanism) so that it can be reproduced from memory after a minimum amount of practice. In such a case it becomes important to test hypotheses about the factors involved in the relation between length and difficulty, and to design procedures which minimize the difficulty of lengthy tasks.

An obvious hypothesis concerning the theoretical factors involved in the above empirical relation is that, as length increases, intra-serial interference increases. With this hypothesis, any procedure which minimizes such interference should reduce the amount of practice necessary to obtain complete mastery. The theoretical factor for reducing interference in the present study is the factor of organization of the task into sub-units, which makes the entire task easier to master by virtue of the greater ease of serially learning a small sub-unit and the ease of serially connecting the much-reduced number of sub-units when compared with the total number of elements in the entire task.

---

[1] This chapter is based on two of the studies conducted under Contract AF 41(657)-45 between Boston University and the U. S. Air Force. The results were previously reported for limited distribution by the authors in two Technical Memoranda (Margolius, Sheffield, and Maccoby, 1957 a, b). Table 1, Chapter 9, shows the relation of these two experiments to those reported in Chapters 7 and 8.

Without qualification, this theoretical approach would appear to reduce to the classical "whole-versus-part" problem described in the human serial-learning literature. This problem poses the question of whether it is easier to learn parts of a task separately and string them together, or to string all of the elements together as a whole. As is well known, this problem has no clear-cut answer; apparently a number of factors combine to make one method superior under one set of conditions and the other method superior under another set of conditions. The present theoretical approach does not propose an answer to this classical problem in favor of part learning. Rather, it adopts the theoretical position that part learning will be favored when the total task can be subdivided into "natural units"—i.e., units having a common context. The theory is that if there is a constant context within units, and a disparate context between units, intra-serial interference can be reduced by any method that separates the learning task into the two parts: (a) serial learning of the order of elements within distinctive units, and (b) serial learning of the order of the distinctive units. In each case the number of items to be serially learned is reduced from the original number of elements, and intra-serial interference is minimized by connecting each element to its own sub-task context. In an ideal case, for example, a 20-item serial problem might be reduced to four five-element units, the four units being within the memory span for units, and the five elements per unit being within the memory span for elements, the entire problem being memorized in a single repetition because both aspects of the total task are within the memory span of the learner.

With these theoretical factors in mind, the purpose of the present study was to determine whether mastery of a serial mechanical-assembly task was facilitated by a practice procedure which enhanced separate mastery of sub-units, as contrasted with a practice procedure which enhanced more the stringing together of the separate elements. The learning task (assembly of an automobile distributor) was one used previously in studies (see Chapter 3) in which various methods of interspersing demonstration and overt practice had been compared. The results of the previous studies indicated that overt practice was more effective if it was interspersed at the completion of demonstration of sub-units of the task than if it were at the end of a complete demonstration of the entire task. The present study was an attempt at a direct test of the hypothesis that learning of the entire task would be facilitated if sub-assemblies were individually better learned before proceeding to a new sub-assembly, as contrasted with an equal amount of practice time in which the learner proceeds to a new sub-assembly after a lesser degree of mastery of each sub-assembly. [In all cases, in this experiment and those reported in the rest of this chapter and in Chapters 7 and 8, demonstration of any sub-unit was—as could be recommended by results of the experiments reported in Chapters 3 and 4—followed immediately by practice of the sub-unit just demonstrated. Thus, whereas the factors previously studied concerned variation in the length of practice sessions and their placement in relation to demonstration sequences, in this chapter (and Chapter 7) the experimental question lay in how demonstration and practice units could best be arranged.]

Specifically, the present experiment compared the case in which each sub-assembly was demonstrated and practiced twice, before any demonstration or practice of the next sub-assembly, versus the case in which each successive sub-assembly was demonstrated and practiced once, but the entire unit-by-unit demonstration and practice sequence was then repeated. The sub-assemblies were selected as natural sub-units of the total task on grounds of common context within each sub-assembly but different context between sub-assemblies.

Method

Demonstration Material

A 16-mm. black and white film 18 minutes in length demonstrated the step-by-step assembly of an automobile distributor.[2] The component parts of the assembly were displayed in the film and their assembly was depicted in sequential order. The narration identified the parts as each was introduced, and described the actions as the part was added into the assembly.

Practice Material

An automobile distributor of the same model as that depicted in the demonstration was used. The component parts were laid out by the experimenter in a standardized pattern on a large white masonite sheet before each practice period. A screwdriver and tweezers were available to S, the subject (student), to utilize in assembly.

Subjects

Twenty male undergraduates at Boston University served as volunteer subjects. Each S was assigned to one of two groups. All S's were paid for the three-hour experimental period.

Experimental Design

All S's viewed the entire film twice and had an opportunity to practice the assembly in conjunction with each demonstration. At the conclusion of training, each S was asked to assemble the distributor without any further film instruction. The third and final assembly was considered as a terminal performance test.

The film was divided into three segments, each considered to be a "natural unit" (see Chapters 3 and 5). The two groups were differentiated with respect to the order in which demonstration and practice of the three segments were scheduled throughout training. Treatments are outlined below; d and p stand for "demonstration" and "practice," respectively, and the subscripts refer to specific segments within the training task:

Group A—"Repetitive": d_1p_1, d_1p_1; d_2p_2, d_2p_2; d_3p_3, d_3p_3; Test.

Group B—"Consecutive": d_1p_1, d_2p_2, d_3p_3; d_1p_1, d_2p_2, d_3p_3; Test.

---

[2]The distributor-assembly film was one adapted by Boston University from a film made by Yale University under an earlier Air Force contract (see Chapter 3).

It will be noted that all groups received the same amount of film demonstration and practice; they differed only with respect to location of practice. Both treatments involved the demonstration of a section and then the practicing of that section by the subject. In the "Repetitive" group, the demonstration and practice were repeated before going on to the next section; in the "Consecutive" group, each demonstration and practice was followed by moving to the next section, this procedure being repeated twice.

Details of Training Procedure

In both groups, S was seated at a table upon which the distributor parts had previously been placed and covered from view. S was instructed to observe and listen carefully to the material presented in the filmed demonstration, since he would soon be called upon to assemble the parts himself. The room light was then lowered and the film projector was started. At the end of d_1 the projector was stopped and the room light was turned up. S was then instructed to select the parts necessary to complete the assembly of the distributor through p_1 as quickly and accurately as possible. When S indicated that he was finished, the experimenter (E) recorded both time and errors, corrected the selection if necessary, and then covered the remaining, or non-selected, parts. S was further instructed to assemble the correct selection of parts as quickly and accurately as possible and to inform E upon completion.[3] While S was at work, E recorded temporary errors (see Chapter 3) and errors made and corrected by S during the practice period. When S indicated completion or exceeded the time limit, E removed the practice material and recorded fixed errors and assembly time. An error was recorded for each uncorrected faulty response or for any omission of a part. Assembly time, taken to the nearest second, did not include the time taken to screw parts together; E stopped his watch whenever S placed the screwdriver in position, and started it again when S removed the screwdriver.

Procedure Specific to Repetitive Group

Upon the completion of p_1, a completely disassembled distributor was placed on the table, replacing the partially completed assembly. The parts were covered, room lights were turned down, and Section 1 of the film was projected once again. This was followed by a second practice of Segment 1. Segments 2 and 3 were run in a similar fashion. Before each period of assembly, E corrected the previous unit, if necessary, before returning the work to S for the necessary additions. Thus, as S was to begin the assembly of Section 2, E handed him a correct assembly of Section 1. S never observed corrections being made, but was informed at the beginning of the next practice unit that the assembly up to that point was now correct.

Procedure Specific to Consecutive (or "Continuous") Group

Upon the completion of $d_1 p_1$, the parts were covered, room lights were turned down, and Section 2 of the film was projected. This was followed by practice of

[3]Complete instructions are given in Margolius, Sheffield and Maccoby, 1957 a.

Section 2, and then demonstration and practice of Section 3. Following the three film and practice periods, the entire procedure was repeated a second time. Table 1 indicates film length, time limits for selection and assembly, and number of correct responses possible for each of the three sections.

TABLE 1

Breakdown of the Task in Terms of Length of Demonstration, Time Limits for Assembly in Practice, and Number of Possible Correct Responses (Task Elements)

Section	Film Length	Time Limits		Correct Assembly Responses
		Selection	Assembly	
	Min:Sec	Min:Sec	Min:Sec	
1. Shaft unit	5:30	2:15	4:30	9
2. Plate unit	6:30	2:30	8:00	11
3. Exterior unit	6:00	2:15	7:30	10
Totals	18:00	7:00	20:00	30

Test Procedure

For the final test, S was presented with the completely disassembled distributor parts and instructed to select the parts and assemble the complete mechanism without any break between units. E recorded assembly time and errors for each of the units separately, as before, although S proceeded without a break.

Results

The present analysis will treat separately performance during the practice trials and performance during the final test.

Practice Findings

The two measures taken, time and correct responses, were combined to give a single meaningful index of performance efficiency. The justification for the single index was that in this study, as well as in the previous study with this task (see Chapter 3), the time scores and error scores showed the same pattern of effects. The index, number of correct responses per unit of time, was obtained by dividing total correct responses by assembly time. The ratio so obtained, called "performance rate," was computed for each S at each practice session. This measure is the same one as was used in the previous research with this task (Chapter 3). Means and variances of performance rates are listed in Table 2, and the data are presented graphically in Figure 1. The superiority of the Repetitive treatment is evident at both stages of practice.

TABLE 2

Means and Variances of Performance Rates
At Successive Stages of Practice

Treatment	1st Practice Mean	1st Practice Variance	2nd Practice Mean	2nd Practice Variance
Repetitive	1.59	0.51	2.75	0.97
Consecutive	1.04	0.13	1.88	0.72
Difference	0.55		0.87	
t	2,07		2.01	
P	<.05 (one tail)		<.05 (one tail)	

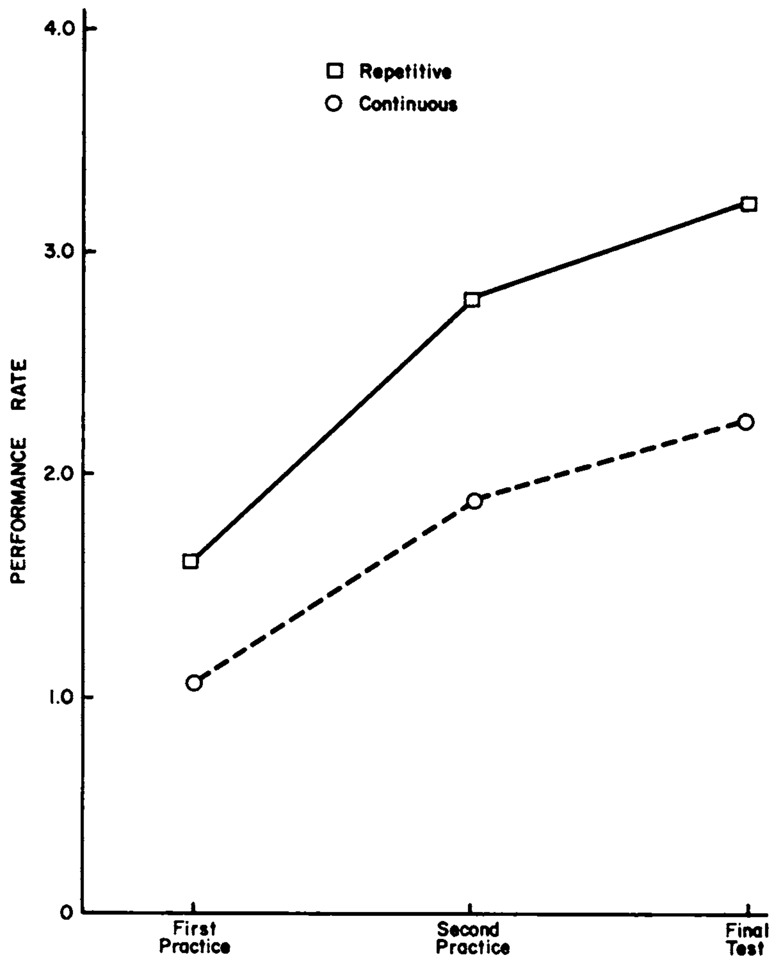


Figure 1. Performance Rate During Practice and Testing (Ignition Distributor)

Final Test Findings

Final test performances and comparisons between the two groups are given in Table 3. The superiority of the Repetitive group is again evident and is in agreement with results obtained during the practice sessions.

TABLE 3

Means, Variances, and Comparison of Performance Rates During Final Test

Treatment	Mean	Variance
Repetitive	3.21	1.35
Consecutive	2.24	1.12
Difference	0.97	
t	1.85	
P	<.05 (one tail)	

Thus, in both practice and test, the method of viewing and practicing a natural unit twice before progressing to the next unit was superior to the method of viewing and practicing such units only once at a time but with two repetitions of the entire demonstration and practice sequence. This result was found not only for the task as a whole, but also for each separate segment of the task. This pattern is shown in Table 4.

TABLE 4

Breakdown of Results into Separate Units of the Task

		Segment 1		Segment 2		Segment 3	
		Mean	Variance	Mean	Variance	Mean	Variance
Practice #1	Repetitive	3.11	1.09	0.79	0.78	2.05	0.74
	Consecutive	2.87	0.82	0.41	0.17	1.00	0.31
	Difference	0.24		0.38		1.05	
Practice #2	Repetitive	4.79	1.01	2.26	1.55	2.71	1.03
	Consecutive	3.24	1.14	1.53	1.29	1.84	0.58
	Difference	1.55		0.73		0.87	
Test	Repetitive	3.88	1.68	3.22	2.27	3.14	1.63
	Consecutive	3.50	1.27	2.25	1.48	1.78	1.28
	Difference	0.38		0.97		1.36	

Interpretation

The results substantiate the hypothesis that the learning of a lengthy serial task can be facilitated when it is broken into "natural" (contextually similar) units, by giving adequate separate practice on each of the individual units before going on to the next unit. Presumably this procedure reduced the total task to three smaller tasks, each of which could be learned as a separate unit whose elements did not interfere with elements in the other units. A related presumption is that the correct order of the three sub-tasks could be learned in a single trial, since the Repetitive group practiced this aspect of the total sequence only once.

The results might also be regarded as favoring the Part method of learning over the Whole method. In such an interpretation it should be clear, however, that the parts were selected as natural units, and that the same outcome would not necessarily be expected if the total task had been subdivided into three parts on some other basis. In this study, the three parts were selected on the basis of common context, and were not merely a subdivision of the entire serial-learning task into successive thirds. It should also be noted that the present method of learning does not exactly fit the classical part-versus-whole type of study. Thus the present method involves an admixture of demonstration and practice, as contrasted with the classical studies based on practice only. To the extent that demonstration involves perceptual learning, and practice after demonstration requires translation of perceptual learning into overt behavior, the present study involves a somewhat different process from that of the rote learning of verbal material by the part-versus-whole method.

We believe, however, that the same principles apply in either case. Thus, if a serial verbal-learning task could be organized into subdivisions of elements having common context, the Part method should prove superior to the Whole method. Similarly, if a poem were written so that each stanza involved a completely different topic, the poem should be easier to learn by the Part (stanza) method than by the Whole method. On the other hand, if the poem merely related a sequence of events, with stanzas being an arbitrary separation of the events, the advantage of the Part method should either be smaller or prove to be a disadvantage when compared to the Whole method.

Part 2: Experiment Using the Waste-Gate Motor Assembly Task

Background and Purpose

The present study was essentially a repetition of the experiment reported in Part 1, but with a different task. The new task—assembly of a waste-gate motor[4]—

[4]The waste-gate motor is a servomechanism used as part of the supercharging system in a B-36 airplane engine. Its sole function is to control exhaust pressure in accordance with intake pressure, but it involves a large number of parts. The actual number of parts in the assembly can be varied depending on how many units are preassembled, or on how much of the complete assembly is required. In the task used, the total assembly was limited to 64 assembly operations—that is, 64 distinct steps in which a part or set of parts is affixed in its correct place in the mechanism.

involved 64 assembly operations as contrasted with the 30 assembly operations of the distributor assembly used previously. Also, the new task was specifically selected from a number of other possible tasks because it was readily subdivisible, on a priori grounds, into distinct sub-assemblies, each of which could be integrated as a sub-unit but which had a distinctive context compared with other sub-units. It was thought that these conditions—lengthy total task but distinctive sub-tasks—would maximize the conditions in which the Repetitive method would be more effective than the Consecutive method in reducing the intra-serial interference. Thus, it was thought that the length per se would maximize intra-serial interference, and that presence of clearly integratable, although lengthy, sub-units would permit minimization of such interference when the Repetitive method was used, as compared with the Consecutive method.

Method

Demonstration Material

The demonstration material consisted of a 16-mm. black and white film which depicted the step-by-step assembly of the 64-part mechanism. The total assembly included a number of identifiable sub-assemblies of varying sizes. For this reason the task could be organized for demonstration and practice into a varying number of separate assembly units. By keeping larger sub-assemblies intact and by combining some smaller sub-assemblies into larger sub-assemblies of which they were parts, the task was divisible into four natural-unit segments. These were as follows:

1. Activator Sub-Assembly
2. Gear Box Sub-Assembly
3. Motor Sub-Assembly
4. Solenoid Brake Sub-Assembly.

In the filmed demonstration the component parts were selected as needed and correctly fitted together in a sequential order. The narration identified the parts and described the appropriate actions as the part was added into the assembly. The film image seen by the subject was 40 x 31 inches, and was approximately 13 feet from S. Good visibility was inherent in the filming, which used adequate close-ups wherever small parts were involved. The film running time for the above sub-assemblies is shown in Table 5 below:

TABLE 5

Film Running Time for Sub-Assemblies

Segment	Time
1 (Activator)	7:59
2 (Gear Box)	4:43
3 (Motor)	6:18
4 (Solenoid Brake)	4:42
Total	23:42

Practice Material

A mechanism of exactly the same model as that depicted in the filmed demonstration was used. The component parts were laid out by E in a standardized pattern on a large white plywood sheet before each practice period. The layout of parts was identical with that used in the film demonstration. Screwdrivers of two sizes and a pair of tweezers were made available to S.

Subjects

Twenty male undergraduates at Boston University volunteered to serve as subjects. Each S was assigned at random to one of the two groups. All S's were paid for the three-hour experimental period.

Design

S's in both groups viewed the entire film twice and had an opportunity to practice each of the four segments of the assembly following the demonstration of that segment. At the conclusion of training, S was asked to complete the assembly without any further film demonstration. The difference in treatment between the two groups was the same as that in the previous study; that is, they were differentiated with respect to the order in which demonstration and practice were scheduled during training. The two treatments in the present instance are outlined in specific terms below, where d and p stand for "demonstration" and "practice" respectively, and where the subscripts refer to the four sub-assemblies within the total task:

Repetitive: d_1p_1, d_1p_1; d_2p_2, d_2p_2; d_3p_3, d_3p_3; d_4p_4, d_4p_4; Test.

Consecutive: d_1p_1, d_2p_2, d_3p_3, d_4p_4; d_1p_1, d_2p_2, d_3p_3, d_4p_4; Test.

Training and Test Procedures

The training and test procedures used were similar to those used in the previous study, with one exception. The selection of parts necessary to complete the assembly of a segment of the present mechanism was not treated as a separate task, but was incorporated as a part of the assembly task. Thus, following each demonstration, S was instructed to select and assemble the parts necessary to complete the unit.[5] This change was made so that performance during training could be compared more directly to performance during the final test, where selection is of necessity an integral part of the task. Table 6 indicates number of correct responses and time limits for assembly for each of the four sections.

[5]Complete instructions are given in Margolius, Sheffield and Maccoby, 1957 b.

TABLE 6

The Task Layout

Segment	Correct Responses	Minutes Allowed for Assembly
1	18	11
2	13	6
3	20	15
4	13	8
Total	64	40

Results

As in the previous study (Part 1 of this chapter), the results are presented here in terms of performance-rate scores. These scores utilize an index which simultaneously takes into account both time to complete the attempted assembly and number of correct responses. Specifically, it is the ratio of number of correct responses per minute of performance. This index is justified on the grounds that both the correct-response data and the time data showed the same general pattern of results. Means and variances of performance rates during practice are shown in Table 7.

TABLE 7

Means and Variances of Performance Rates
At Successive Stages of Practice

Treatment	1st Practice		2nd Practice	
	Mean	Variance	Mean	Variance
Repetitive	1.70	0.60	4.26	1.09
Consecutive	1.65	0.78	2.94	0.56
Difference	0.05		1.32	
t	--		3.07	
P	--		$<.01$	

These results agree with the previous findings to the extent that the Repetitive group was superior to the Consecutive group during practice. This difference was not particularly in evidence during the first practice, but emerged as a statistically significant difference in the second practice.

The most important discrepancy between the previous results and those of the present study is that in the previous study (Part 1 of this chapter) the superiority of the Repetitive group was in evidence not only in practice, but also in the final test of the previous study. In the present study, on the other hand, the superiority achieved by the Repetitive group during practice disappeared in the final test. This result is in evidence in Table 8, which shows means and variances of performance

TABLE 8

Means and Variances of Performance Rates During the Final Test

	Mean	Variance
Repetitive	4.13	1.41
Consecutive	4.05	1.51
Difference	0.08	

rates during the final test. It also can be seen in Figure 2, which depicts practice and test data in the form of learning curves. The figure brings out clearly the fact that, whereas the Consecutive (or "Continuous") group continued to improve from second practice to test, the Repetitive group dropped slightly in the transition from the demonstration-and-practice conditions to the conditions of final testing. The slight drop from practice to test in the Repetitive group is not at all significant, but the rise in the Consecutive group from practice to test is highly significant ($t = 5.13$, $p<.01$), and the differential effect between the two groups is also very significant ($t = 2.952$, $p<.01$). That is, the slight drop in the Repetitive scores (4.13 - 4.26) differs significantly from the sizable rise (4.05 - 2.94) in the Consecutive scores.

Because they may be of some significance in the interpretation of the above findings, the means of both the time scores and the correct-response scores (from which the performance rates were derived) are presented in Table 9. The most notable thing in the table is that the mean number of correct responses is at a very high level in both the second practice and the final test for both the Repetitive and the Consecutive groups. The total possible number of correct responses was 64, and the number obtained in both second practice and test is over 90 per cent in either group in each case. It should be noted that the performance-rate means already shown cannot be exactly derived from Table 9, since mean performance rates were obtained by averaging each individual's performance rate rather than by dividing average number correct by average time.

Interpretation

The results of this study are clearly contrary to the initial expectation. That is, the task was selected to maximize the likely advantage of the Repetitive over the Consecutive method of demonstration and practice, yet the results showed superiority for the Repetitive method only during the second practice, and this superiority was lost in the test.

One obvious factor that may be relevant is that, despite the length of the serial task (64 elements to be assembled), the training was so effective that after only two

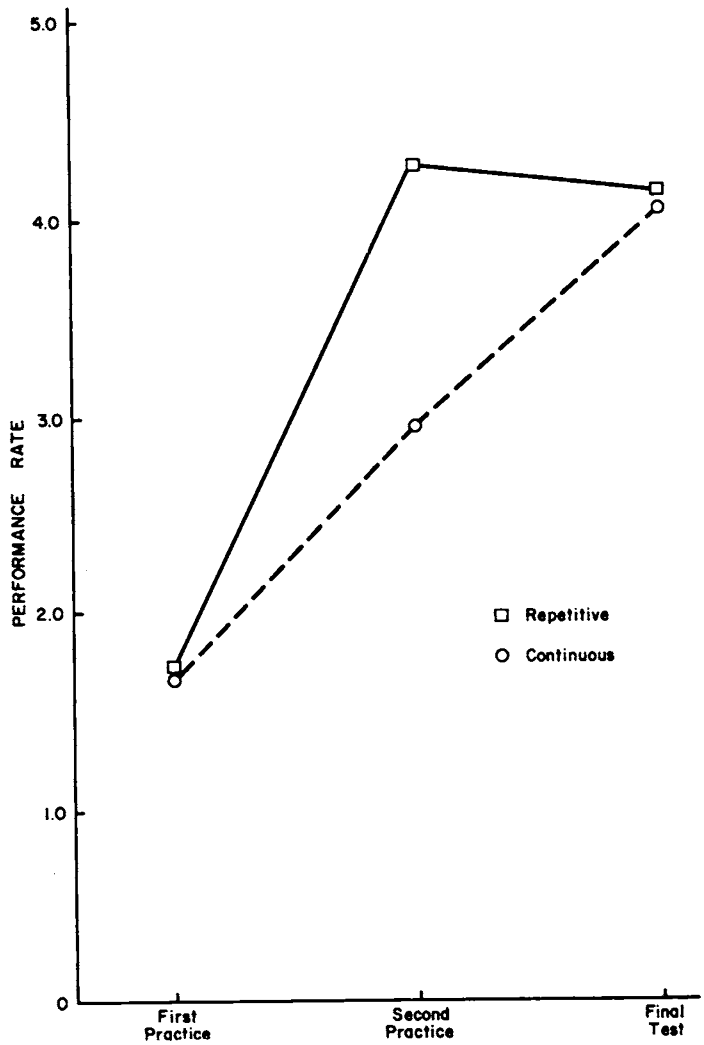


Figure 2. Performance Rates During Practice and Test (Waste-Gate Motor)

TABLE 9

Mean Time and Mean Correct Responses During Practices and Test

	1st Practice	2nd Practice	Test
Time			
Repetitive	27.8	15.2	15.4
Consecutive	33.0	21.4	15.6
Correct Responses (Total Correct Responses Possible = 64)			
Repetitive	46.6	62.6	59.6
Consecutive	49.3	58.7	59.1

demonstrations and overt practices of the sub-assemblies, the entire task could be almost perfectly assembled in the final test. Thus, intra-serial interference may have been at a minimum for other reasons, making relatively ineffective any procedure designed to reduce the interference that might otherwise be expected.

In support of this interpretation is the fact that final test performances showed over 92 per cent correct responding in both groups (Table 9). Also in support of this interpretation is the fact that the film was prepared so as to maximize advantages of an organized presentation that would minimize interference. This was done in connection with another purpose not strictly relevant to the present report. It becomes relevant in the present context only because it was thought that, despite the well-organized presentation, the difficulty of serially learning sub-assemblies with 18 (Segment #1) and 10 (Segment #2) elements would be sufficiently great that an advantage would be achieved by the Repetitive as compared with the Consecutive method.

It should be noted in this connection that an earlier study (see Chapter 3) indicated that practice interspersed at the end of contextually similar units reduced intra-serial interference as contrasted with equal practice at the end of a demonstration of the complete task. Since this interspersed-practice procedure was used even in the Consecutive method of the present study, there may have been very little further interference left to reduce by the Repetitive method. If this were the case, then the Consecutive method would lose little in terms of interference between elements in different sub-units, and would gain something in terms of integration of separate units. The Repetitive procedure, on the other hand, would gain something in the reduction of interference between elements of different units, but would lose something in terms of the integration of separate units.

Thus, it may be that a task which is intrinsically well organized, and which is presented with due attention to this organization, will be learned readily with little intra-serial interference. In such cases acquisition cannot be greatly improved by procedures designed to minimize interference between elements of the separate subtasks. By the same token, in such cases it may be desirable to concentrate more on the serial connecting of the larger sub-units of the task, than on the integration of the elements within a sub-task. In the present instance, the outcome may be interpreted as a "tie" between these two factors if only the final test is considered. This "tie" in the final test, however, was accompanied by a significant differential improvement between the two groups from second practice to test, the Consecutive group improving more although starting from a lower value. Presumably the Consecutive group was penalized by greater interference during transition, but the Repetitive group was penalized by lack of total task integration in the test.

Evidently in the previous study (see Part I), with only three rather than four natural units and without organizability of sub-tasks as a factor in the choice of task, the serial connecting of sub-tasks was a less important problem, and the integration of elements within sub-tasks was a more important problem. Under these circumstances the Repetitive procedure was superior.

CHAPTER 7

TIMING OF DEMONSTRATION AND OVERT PRACTICE AS A FUNCTION OF TASK ORGANIZATION[1]

Garry J. Margolius
Fred D. Sheffield
Nathan Maccoby

Background and Purpose

The purpose of the present study was to test the hypothesis that, in learning a lengthy mechanical-assembly task by joint use of demonstration and overt practice, the optimum timing of overt practice is at the completion of each natural unit of the total assembly task. The meaning of "natural unit" in this hypothesis—as well as the general theory giving rise to the hypothesis—has been discussed in Chapter 2. Briefly, a natural unit of a task is a portion of the sequence which has its own distinctive context cues, and the general theory implies that if such portions are practiced as separate units, their elements will be less subject to interference from similar elements in different units. More specifically, the theory says in part that if a lengthy serial task can be segregated into several sub-tasks, the serial learning of the separate sub-tasks is simplified by virtue of the smaller number of units to learn. It also says in part that if individual elements are practiced within the distinctive context of their own sub-task, they will tend to be elicited only in that context and therefore will not compete with similar elements in other sub-tasks.

The particular method of testing these implications in the present study consisted of interspersing overt practice either (a) at the end of *a priori* natural units or (b) at the end of arbitrarily determined segments of the task which matched the natural units in their temporal length. The former method will be referred to as "Synchronous" method because practice was synchronized with the termination of demonstration of each natural unit of the task. The other method is called "Asynchronous" because the practices come at other times during the demonstration. It should be clear. however, that the same total amount of practice and the same number of practice periods were used in each case.

[1]This chapter is based on one of the studies conducted under Contract AF 41(657)-45 between Boston University and the U. S. Air Force. The results were previously reported for limited distribution by the authors in a Technical Memorandum (Margolius, Sheffield, and Maccoby, 1957 c). Table 1, Chapter 9, shows the relation of this study to those reported in Chapters 6 and 8.

Method

The task was the assembly of a waste-gate motor, a task involving 64 assembly operations. This task can be divided on a priori grounds into four sub-assemblies, each having a distinctive context in which its assembly operations are carried out (see Chapter 6). The method of practice for the Synchronous groups was to stop a filmed demonstration at the end of each unit, practice that unit, and then re-demonstrate the same unit and practice it once more before proceeding to the next unit. The method with the Asynchronous group was identical in every respect, except that the stopping points for practice came at different times which were out of phase with the completion of the four sub-assemblies. The exact nature of the difference in treatment is illustrated in Figure 1. It can be seen in the figure that the lengths of the time segments were matched between the groups, the only difference being the location in time of a given sized segment.

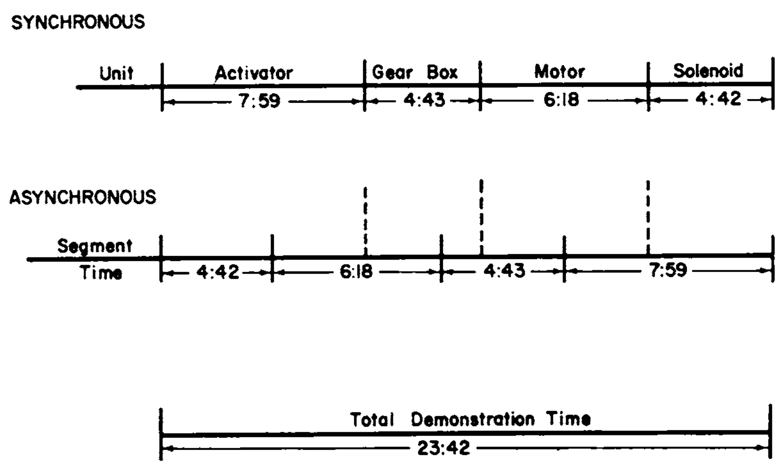


Figure 1. Distribution of Demonstration Time Segments with Synchronous and Asynchronous Treatments. Time is Shown in Minutes and Seconds; the Dotted Line in the Asynchronous Group Corresponds with the Solid Line in the Synchronous Group and Shows Where Natural Units of the Task Terminated in the Demonstration.

The number of correct responses in each segment of the demonstration-and-practice sequence, and the time limits allowed for completion of a given practice segment, are shown in Table 1.

The filmed demonstration, the practice materials, and the details of the practice and testing procedure were identical with those described in Part 2 of Chapter 6. In fact, the present Synchronous group corresponds exactly in all features of procedure with the Repetitive group in the previous experiment. A new set of subjects was run for the Synchronous group in the present study only because of the possibility of differences in the populations sampled, in view of the fact that the two studies were carried out at different times. The present experiment was carried out on 20 volunteer undergraduates at Boston University. S's were paid for their services.

TABLE 1

Length of Segment, Number of Correct Responses, and Time Limits Within the Two Treatments

Segment	Synchronous			Asynchronous		
	Demon-stration Time	Correct Responses	Minutes Allowed to Practice	Demon-stration Time	Correct Assembly Responses	Minutes Allowed to Practice
1st	7:59	18	11	4:42	10	6
2nd	4:43	13	6	6:18	12	8
3d	6:18	20	15	4:43	18	11
4th	4:42	13	8	7:59	24	15
Total	23:42	64	40	23:42	64	40

Results

The results showed no effect of the difference in treatment. The Synchronous group was slightly faster in time scores, but the Asynchronous group was slightly better in number of correct responses; however, none of the differences is significant at either stage of practice or in the final test given at the end of practice. Means, variances, and *t*-tests are shown for time scores in Table 2 and for correct response scores in Table 3. The net result of the slight advantage in time scores for the Synchronous group and the slight advantage in correct response scores for the Asynchronous group was almost identical performance rate scores in the two groups. This outcome is shown in Figure 2 as the ratio of number of correct responses per minute of performance.

TABLE 2

Means, Variances, and *t*-tests for Time Scores at Each Stage of Practice and Test

	1st Practice		2nd Practice		Test	
	Mean	Variance	Mean	Variance	Mean	Variance
Synchronous	30.4	43.0	19.2	66.8	20.5	56.4
Asynchronous	31.8	42.0	21.3	115.6	22.9	123.6
Difference	1.4		2.1		2.4	
t	0.47		0.48		0.55	

TABLE 3

Means, Variances, and t-tests for Correct Response Scores at Each Stage of Practice and Test

	1st Practice		2nd Practice		Test	
	Mean	Variance	Mean	Variance	Mean	Variance
Synchronous	39.8	120.4	56.0	65.6	51.4	119.0
Asynchronous	45.0	182.0	59.3	36.4	53.8	117.6
Difference	5.2		3.3		2.4	
t	0.89		0.97		0.40	

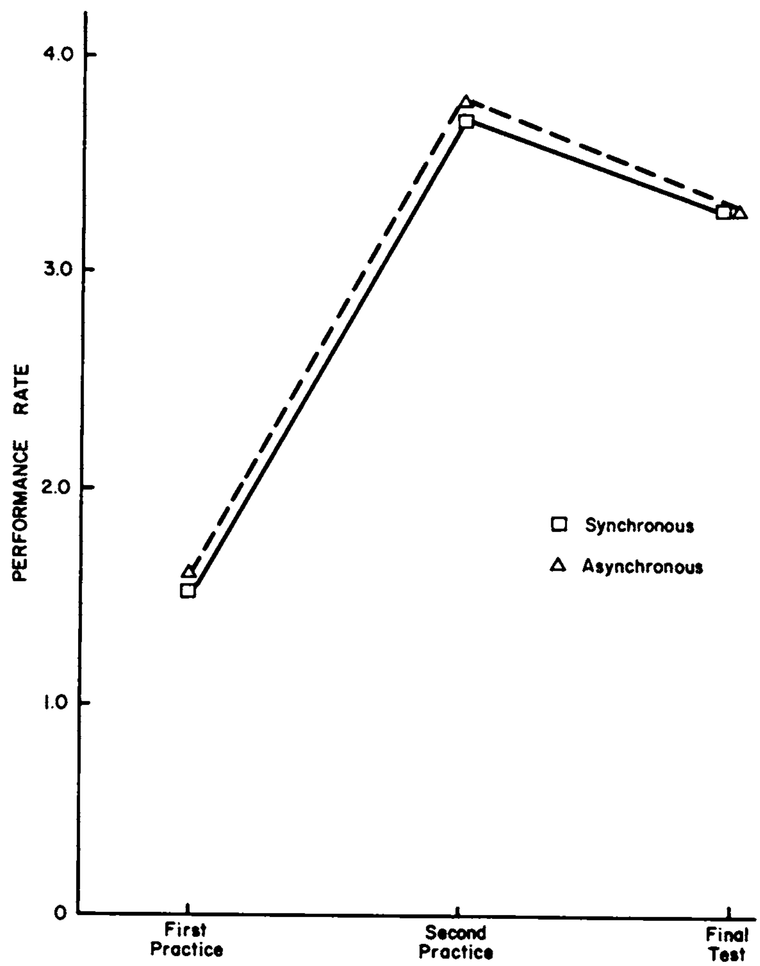


Figure 2. A Comparison of Performance Rate Scores During First Practice, Second Practice and Final Test

Interpretation

One possible implication of these results is that if a task is inherently well-organized, there is very little to gain from additional procedures designed to enhance the organization. This conclusion is contrary to the original theoretical expectation, the expectation being that if a task is inherently easy to organize, any procedure which enhances this organizability would enhance the ease of learning. The results, however, show that in each study, with the same task employed, separate methods designed to enhance the integration within natural units added nothing to performance in the final test. At the same time, another study using a different task (see Part 1 of Chapter 6) showed a clear-cut gain from a procedure designed to enhance integration within natural units. In this other study the task had not been preselected for inherent organizability. This fact tends to strengthen the above possible conclusion that gains through methods capitalizing on the integration of natural units will be small if these units tend to be easily integrated by the learner without special methods.

In connection with this interpretation it should be pointed out that in both studies the method of increasing the advantage through organization of natural units involved some disadvantages in other respects. Thus, in the previous study, the enhancement of learning within natural units necessarily involved a sacrifice in the integration of the separate units. In the present study, this sort of sacrifice may also have been a factor. Thus, the Asynchronous group viewed the demonstration and practiced performance across natural units. This can be seen clearly by referring back to Figure 1, which shows the design. It can be seen in this figure that the Asynchronous group had experience with the <u>transition</u> from each natural unit to the next one, whereas for the Synchronous group such junctures were always stopping points, with no continuity, either in demonstration or in practice. To this extent the Asynchronous group would be expected (like the Consecutive group of the previous study) to gain some of the advantages of the Whole method of learning, which emphasizes integration of the entire task. (See also further discussion in Chapter 9.)

CHAPTER 8

EXPERIMENTS ON PERCEPTUAL MEDIATION IN THE LEARNING OF ORGANIZABLE SEQUENCES[1]

Fred D. Sheffield
Garry J. Margolius
Arthur J. Hoehn

The experiments reported in the preceding chapters have dealt with methods of utilizing an already prepared demonstration. The present experiments deal with methods of constructing a demonstration, although utilization variables were also involved. The experiments were planned in terms of organizational factors as discussed in Chapter 2, with particular emphasis on the "perceptual-blueprinting" concept.

The serial-learning task employed in the experiments was the mechanical assembly of the waste-gate motor, described in Part 2 of Chapter 6. As noted in Chapter 6, the assembly task was one which could be divided into four sub-assemblies: (a) the "activator" assembly, (b) the "gear box" assembly, (c) the "motor" assembly, and (d) the "solenoid brake" assembly. Even these sub-assemblies could in some cases be broken into still smaller sub-assemblies,[2] but the demonstration, which was a narrated film, was organized for presentation into the four assemblies named above.

The plan of the present experiments was to study the effects of formation of "perceptual blueprints" by exposing each completed sub-assembly at the end of each unit of the assembly task. Thus the expectation is that if the learner had an opportunity to examine visually each completed unit, he would acquire a perceptual image of the correct location of parts which would guide his future attempts at assembly. An inherent difficulty with this technique in the present task, and perhaps in most mechanical-assembly tasks, is that the completion of each unit also conceals the inner works in the housing of the mechanism. Thus there is no good way of exposing the total sub-unit without recourse to such artificial methods as plexiglas housings and parts, "cutaway" models, and so forth.

[1]This chapter is based on three experiments conducted under Contract AF 41(657)-45 between Boston University and the U.S. Air Force. The results were previously reported for limited distribution in a Technical Memorandum (Sheffield, 1957a). Table 1, Chapter 9, shows the relation of these experiments to those reported in Chapters 6 and 7.

[2]E.g., one of the "gears" in the gear box assembly had nine parts to assemble before the gear was put in place.

The solution devised in the present instance was a technique termed "implosion" of the parts at the end of each sub-assembly. The technique would not be possible with a live demonstration but is easily arranged with a filmed demonstration. It consists of showing a display of the parts of a given sub-assembly and then, through a rapid succession of "stills," showing each part "jump" into its proper place in the proper sequence. This procedure utilizes the "Phi phenomenon" to create the illusion of motion of the parts from the display to their appropriate location in the assembly, and it permitted the condensation of a large number of steps into a brief period of time. The speed used, however, was not the maximum possible because the implosion sequences were also narrated in a condensation of the original narration of the full demonstration of the sub-assembly. The narration was very rapid, however, often consisting of no more than naming the particular part as it jumped into its place. Thus the implosion sequences went very rapidly, the longest one using only 87 seconds for a sub-assembly involving 18 separate assembly operations.

The name, "implosion," for this technique was suggested by the "exploded views" of parts commonly used to show assemblies in technical manuals. Thus the "implosion" method starts with an "exploded" view, and "implodes" the parts into place. From the standpoint of the perceptual theory discussed in Chapter 2, the implosion technique is an attempt to give a "perceptual blueprint" under circumstances in which a rapid sequence must be used rather than a static view of the completed sub-assembly. As an instructional device it should work best if there is still a fairly clear perceptual image of the first part at the time that the last part goes into place. Speed is therefore desirable, but excessive speed would create a chaotic effect because there would not be sufficient time for a full perceptual response to each part. It should be emphasized, however, that the technique was not an attempt to be a form of rapid serial learning. Greater speeds of sequential responding are clearly possible with perceptual behavior than with overt behavior, and it may be that such sequences can be learned more readily, both because of lesser interference among perceptual responses and because of the greater opportunity for organizing a <u>temporally rapid</u> sequence. But in the present instance the attempt was only to provide a static blueprint which would guide the performance by perceptual memory of the structural relation of the parts after they are in place.

Method

As already noted, the filmed demonstration depicted the assembly of a wastegate motor. This assembly task was used in two previous experiments investigating various methods of combining demonstrations with practice (see Chapters 6 and 7). All of the previous reports used the filmed demonstration without the addition of the implosion sequences. The experimental variable in the present report consisted of presence or absence of the implosion sequences, and in all cases one of the treatments in the previous reports supplied the control on the effects of adding the implosion variable. Thus the present experiments actually consisted of the comparison of two different versions of the same filmed demonstration. The difference between the two versions was that in one version the demonstration of each of the four sub-assemblies was followed by an approximately one-minute review of the sub-assembly in the form of implosion sequences.[3] The film-projection time of the sub-assemblies with and without the implosion sequences are shown in Table 1.

TABLE 1

Projection Time in Minutes and Seconds of the Control and Experimental Versions of the Filmed Demonstration[4]

Sub-assembly	Control	Experimental	Implosion Time
1. Activator	7:59	9:55	1:27
2. Gear-box	4:43	6:08	1:03
3. Motor	6:18	8:09	1:20
4. Solenoid-brake	4:42	6:11	1:05

Two separate experiments were run in which comparisons were made between the implosion version and the non-implosion version of the film. In one experiment the implosion method was tested in interaction with a utilization variable described previously as the "Repetitive" method versus the "Consecutive" method (see Chapter 6). Briefly, the Repetitive method consisted of demonstrating and overtly practicing each sub-assembly twice before moving to the next sub-assembly, whereas the Consecutive method consisted of demonstrating and practicing each sub-assembly only once before moving to the next sub-assembly, the entire demonstration and practice sequence being repeated twice. In the other experiment, another utilization variable was used, but this was a three-group experiment in which the implosion factor was not co-varied with the other variable (see Chapter 7). Two of the three groups, however, used the above Repetitive method of demonstration and practice, the only difference being the presence or absence of the implosion technique. Thus the total experimentation relevant to the implosion variable can be outlined as in Table 2.

One reason for presenting the design in the above form is that there was no particular advance reason for expecting any interaction between the implosion variable and the Repetitive-Consecutive variable, and certainly no reason to expect interaction between the Repetitive variable and the "variable" of the difference between Experiment II and Experiment III. The latter difference involved only a slight difference in experimental personnel and a possible difference in the sampling of the undergraduate students who served as subjects at the two different times in the two different experiments. The above design therefore involves, in effect, three different replications of the relevant experimental variable (i.e., implosion) under varied conditions of experimenters, utilization, and population sampled.

[3]The "implosion" version of the demonstration film also included at the beginning of each sub-assembly a shot demonstrating how that sub-assembly fitted into the housing of the waste-gate motor and the main component of the sub-assembly.

[4]Control time plus experimental time exceeds "implosion time" for the reason indicated in footnote 3, _supra_. See also the discussion of the time variable at the end of this chapter.

TABLE 2

Design of Experiments Comparing the Implosion and Non-implosion Versions of the Demonstration

		Treatment	
Experiment*	External Condition	Non-implosion	Implosion
II-A	Consecutive	N = 10	N = 10
II-B	Repetitive	N = 10	N = 10
III	Repetitive	N = 10	N = 10

*Cf. Table 1 of Chapter 9.

Measurements were obtained from the subjects at three different points in performance: (a) in overt practice after the first demonstration of a given sub-assembly, (b) in practice after the second demonstration of a given sub-assembly, and (c) in performance on the final test at the end of training, when the entire assembly was attempted without interruption for any further demonstration. The most important outcome from a practical standpoint is the performance during the final test. This performance can be scored in terms of number of correct-assembly responses, time to complete the assembly, or in terms of a joint measure—"performance rate"—which gives number of correct-assembly responses per unit time in minutes.

Results

As in previous reports on this learning task (see Chapters 6 and 7), the present results will be presented primarily in terms of performance-rate scores. Using this measure, a complete set of means and variances for the six groups, broken down by first practice, second practice, and final test, is shown in Table 3.[5] An overall analysis of variance of this entire table is highly significant, but this fact is not of much importance *per se*. For further analysis the table was collapsed in various ways to bring out the findings that are relevant to the present report.

One obvious fact that can be seen in the table is that all groups showed sizable improvements from first practice to second practice. This result was to be expected for obvious reasons and will receive no further comment, except to note that this improvement in performance-rate is produced both by a significant increase in correct responses and by a significant decrease in time to assemble. Also the mean performance-rate scores do not reveal the fact that the mean correct-response scores were very high by the time of the second practice, most of the groups averaging about 90 per cent or better at this point.[6] This high level of correct responding was not anticipated in such a lengthy task (64 assembly operations) and may have attenuated the effects of the experimental treatment.

[5]Comparable tables have been presented separately for correct-response and time scores in a Technical Memorandum by Sheffield (1957a).

[6]The maximum possible correct responses were 64, and all but one of the groups averaged better than 58 correct responses in the second practice.

TABLE 3

Breakdown of Means and Variances of Performance-Rate Scores by First Practice, Second Practice, and Final Test

Stage of Training		Non-Implosion 1st Prac	2nd Prac	Test	Implosion 1st Prac	2nd Prac	Test
Exp. II-A Consec-utive	Mean	1.65	2.94	4.05	1.99	4.16	4.79
	Variance	0.78	0.56	1.51	1.11	2.03	3.32
Exp. II-B Repet-itive	Mean	1.70	4.26	4.13	2.05	4.43	3.67
	Variance	0.60	1.09	1.41	0.52	1.60	1.33
Exp. III Repet-itive	Mean	1.47	3.68	3.30	2.43	5.93	5.49
	Variance	0.65	3.78	4.27	0.67	2.70	1.53

It also can be noted in the table that each implosion treatment is superior to its corresponding non-implosion treatment at all stages of practice and testing except in one case—the final test in the repetitive group of the first experiment. The overall means (all three "replications") are shown in Table 4, in which it can be seen that despite the above-mentioned exception, the mean performance-rate for implosion was greater than that for non-implosion at both stages of training and in the final test.

TABLE 4

Overall Mean Performance-Rate, for the Three Groups Combined, at the Two Stages of Practice and in the Final Test

	1st Practice	2nd Practice	Final Test
Implosion	2.16	4.84	4.65
Non-implosion	1.61	3.63	3.83
Difference	0.55	1.21	0.82

Separate analyses of variance were carried out for each of the above stages of practice and testing. In the first practice the effects of treatment were significant ($p < .05$), whereas neither the "replication" variable nor the interaction between treatment and "replication" was significant. In the analysis of variance of the second practice the treatment was again significant, and, as in the first practice, the interaction between treatment and "replication" was not significant.[7] That is, the gain

[7] The details of this analysis for the two stages of practice have been presented in supplementary tables by Sheffield (1957a).

from implosion held for all three of the "replications" as broken down by Experiment II and Experiment III, and by Consecutive versus Repetitive conditions in Experiment II. By contrast with the first practice, however, the "replication" variance was significant in the second practice ($p < .05$). This "replication" effect comes primarily from the fact that the Repetitive condition was superior to the Consecutive condition in the second practice. This outcome was reported elsewhere for those groups in which a direct comparison of these two conditions was made (see Chapter 6). The outcome holds in the present data both for implosion and non-implosion and for the comparison of Consecutive groups with Repetitive groups in both Experiment II and Experiment III.

The analysis of variance of the final test data is of greater consequence than the above two analyses concerned with the first and second practice. This greater consequence derives partly from the fact that the performance on the final test is generally of more practical significance than performance during practice, but it also derives partly from the fact that the only exception in Table 3 to the general finding of advantage to implosion was in the final test results. The analysis of variance of the final test data revealed a significant effect of the Implosion treatment despite this exception, but also revealed a significant interaction between treatment and "replication."[8] This significant interaction is to be attributed to the one exception to the general trend of greater mean performance rate under the Implosion treatment. The cell means (taken from Table 3) and the marginal means for this analysis of variance are shown in Table 5, and the analysis of variance breakdown is shown in Table 6.

TABLE 5

Cell Means and Marginal Means for the Analysis of Variance of Performance Rates in the Final Test

Replication/Treatment	Non-Implosion	Implosion	Replication Means
Exp II-A: Consecutive	4.05	4.79	4.42
Exp II-B: Repetitive	4.13	3.67	3.90
Exp III: Repetitive	3.30	5.49	4.39
Treatment Means	3.83	4.65	

TABLE 6

Analysis of Variance of Performance Rates in the Final Test

Factor	S Squares	df	Mean Square	F	P
Treatment	10.21	1	10.21	4.15	<.05
Replication	3.41	2	1.71	0.70	--
Treatment x Replication	17.65	2	8.83	3.59	<.05
Within Cell	132.88	54	2.46		

[8]There was no effect of "replication" *per se* in the final test.

It is apparent from an examination of Table 5 that the exception which accounts for the significant interaction comes from the Repetitive groups in Experiment II. The direction of results from this group is not only opposite to that found for the Consecutive groups in Experiment II but also is opposite to that found for the Repetitive groups in Experiment III, which were given similar treatment. If the interaction had been derived from the comparison of the Consecutive "replication" with the two Repetitive "replications," one might think the Implosion method was more effective under one of these conditions than the other. But the two Repetitive replications combined give results very similar to those of the Consecutive group, as shown in the set of means below:

Mean Performance Rate

	Non-Implosion	Implosion
Consecutive	4.0	4.8
Repetitive (Exp. II plus Exp. III)	3.7	4.6

The significant interaction lies instead in the difference between the Repetitive groups in the two experiments, one of which gave a larger effect of implosion than that formed in the Consecutive group and the other of which gave a reverse effect (although this negative difference was not significant: $t = 0.84$, $p > .10$).

The interaction between the Repetitive groups, which were treated identically and differed only in which set was run first, was quite significant ($F = 7.42$, $P < .01$) from a statistical standpoint but has no significance from a scientific standpoint. That is, an interaction between pure replication and treatment can be interpreted simply as a rare random outcome.

The only interpretable result of the analysis of variance of test scores, therefore, is the significant effect of treatment. Fortunately, this effect met the five per cent level despite the "exception." One may guess, therefore, that the "exception" was accidental and that the true effect of implosion could be better estimated from the two replications which are in line with the overall effects, namely the Consecutive groups of Experiment I and the Repetitive groups of Experiment II. If these four groups are considered by themselves, the effect of implosion becomes larger and more significant and the interaction between "replication" and effect of implosion disappears.

The main findings are, accordingly, presented as a function of treatment for the combined groups in Figures 1, 2, and 3 for performance rates, numbers of correct responses, and time.

Interpretation

The implications of the present experiment are that the implosion technique gives a definite advantage in the effectiveness of a demonstration of a mechanical assembly. In terms of performance on the final test, the mean performance rate was 3.83 in the three control groups and 4.65 in the three experimental groups. If

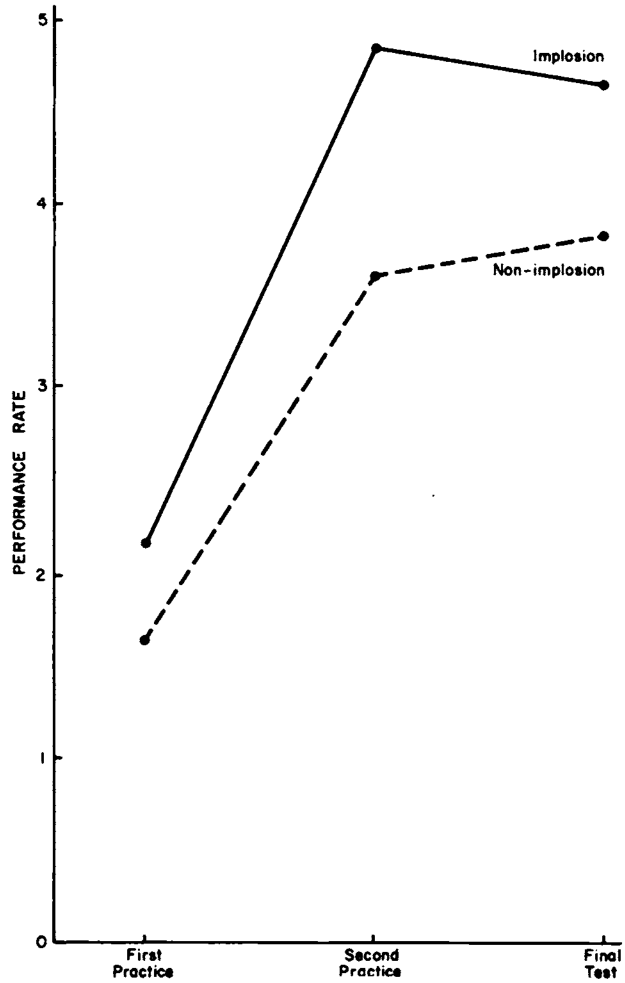


Figure 1. Performance Rates in Completing the Assembly in the Two Practices and the Final Test

the final performance rate in the control groups is taken as a baseline, therefore, the advantage produced by the use of the implosion device was a 21 per cent improvement.

It should be pointed out that in the present results the utilization factors varied somewhat but were held constant between treatments. Thus, two of the comparisons involved the Repetitive method of practice whereas one of the comparisons involved the Consecutive method of practice, but in all cases an Implosion group was compared with a Non-Implosion group under the same external conditions. The fact that the Implosion method came out generally superior is important in view of the fact that some of the previous studies indicated no training advantage to methods calculated to produce organization of sub-assemblies. In these studies of utilization factors, however, training advantages from one standpoint were sacrificed to produce training advantages from another standpoint. Specifically, the training

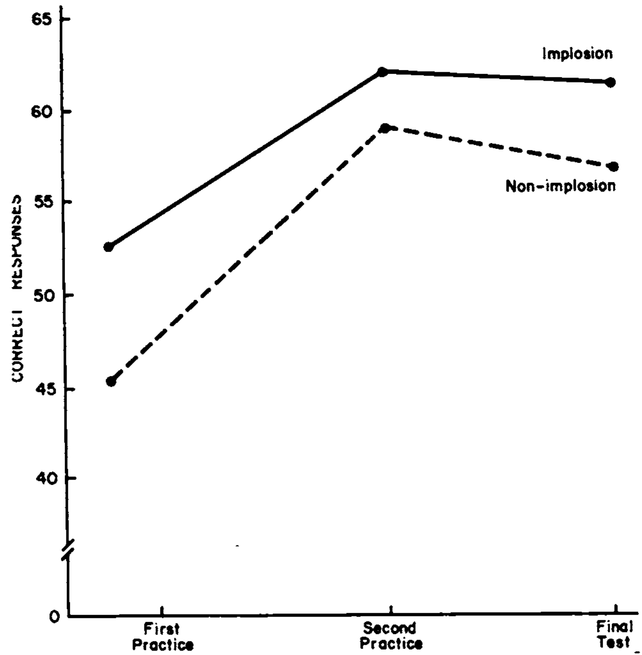


Figure 2. Number of Correct Responses in Completing the Assembly in the Two Practices and the Final Test.

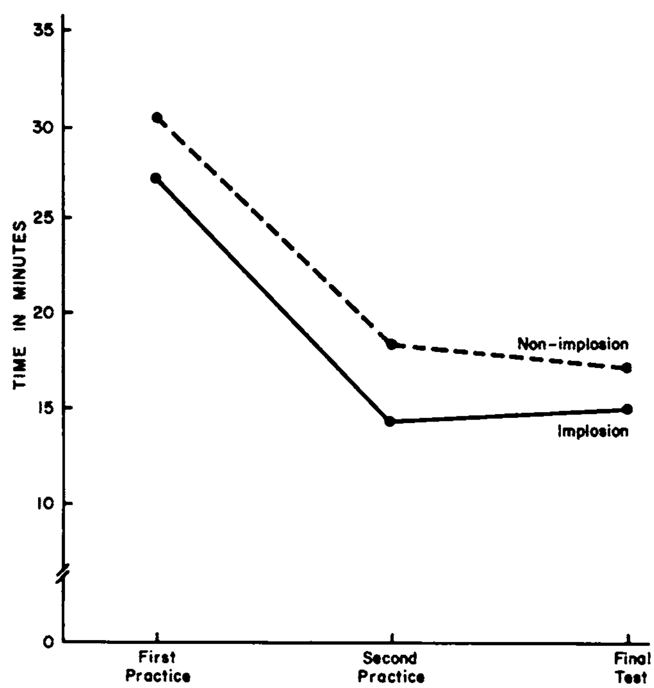


Figure 3. Total Time to Complete the Assembly in the Two Practices and the Final Test.

advantage of organizing the entire sequence was sacrificed in order to capitalize on the training advantage or organization of the separate sub-assemblies. The results of these experiments indicated no difference, which provides no proof that the experimental variables designed to facilitate organization within sub-assemblies had any effect. The present experiments, on the other hand, held constant the utilization factors while varying a factor calculated to enhance the organization of sub-assemblies. The results indicate that the implosion factor, which was designed to enhance the organization of sub-assemblies, did have its calculated effect. One possible conclusion, therefore, is that the other studies also conformed to the general theory but used an unfortunate combination of antithetical effects.

It should also be pointed out that the experimental design does not include any "control" comparison with other filmed material that might have been added to the non-implosion demonstration. That is, the only difference between the implosion and non-implosion films was the presence or absence of the implosion sequences, including the shots of the completed sub-assembly and where it fits into the motor. A necessarily correlated factor was therefore the additional length of the implosion version of the demonstration (as is shown in Table 1), and the question might be raised as to whether the gain from implosion was merely due to the additional time spent in demonstration.

Obviously there is no logical control condition to answer this kind of question, since the additional time could be filled with any one of a large variety of different kinds of relevant filmed material which used equal time. To select one, or several, would essentially be providing a comparison of the implosion technique with some alternative form or forms of review or supplementary material, which was not within the scope of the experiment. Among such various alternatives, the implosion technique seems a very promising one because it provides a complete audio-visual review of an assembly or sub-assembly in a minimum of time and in a manner calculated to maximize organizational factors.

CHAPTER 9

SUMMARY AND INTERPRETATION OF RESEARCH ON ORGANIZATIONAL PRINCIPLES IN CONSTRUCTING FILMED DEMONSTRATIONS[1]

Fred D. Sheffield
Nathan Maccoby

Introduction

As was noted in Chapter 2, the present chapter is an overall summary and interpretation of the studies reviewed in Chapters 6, 7 and 8. The chapter was prepared partly to provide in one place what might be considered a statement of the substance of the research, without going into details. It also attempts to bring out factors the experimenters consider relevant in explaining apparent inconsistencies. It further reflects the judgment of the researchers on how to construct an effective visual aid for teaching complex sequences. This judgment is based partly on theory and partly on the experience accumulated in the studies reported in the chapters of Part I.

Purposes

The basic purpose of the research was to investigate the factor of organization as an aid to serial learning from a combination of demonstration and practice. All the research was specific to the serial learning of a mechanical assembly task by means of narrated motion-picture demonstrations of the task together with overt practice at assembly following demonstration. The organizational principles, however, are considered common to any case of complex sequential learning, whether it be at the simple level of motor coordination or at the complex symbolic level of memorizing a speech.

The present research was also aimed primarily at variables which could be built into the demonstration as contrasted with utilization variables (in which an attempt is made to maximize the training effects of an existing demonstration through appropriate utilization). Both kinds of variables—utilization variables and variables in the demonstration itself—were actually studied, but the chief research effort was on the latter type, i.e., variables concerned with effectiveness of demonstrations regardless of methods of use.

[1]This chapter summarizes the conclusions and interpretations derived from the work under Air Force contract AF 41(657)-45, conducted at Boston University in 1956-57 and reported in Chapters 6, 7, and 8.

To this end a new film demonstration was constructed and narrated. Attention was given to organization in the selection of the task, in the production of the film, and in the narration of the film. An optimum demonstration was produced from the standpoint of organization principles (that is, a poorly organized version was not prepared). The controlled variation within the demonstration consisted of adding to this optimum demonstration further devices designed to enhance the already-present degree of organization.

The utilization variables studied were related to the previous research (Chapters 3, 4, and 5), in which various methods of interspersing overt practice and demonstration had been tested. One of the questions along these lines (considered in Chapter 6) was: If a task breaks down into natural units, is there a mnemonic advantage to repetition of demonstration and practice within a given unit before proceeding to the next unit—as compared with consecutive demonstration and practice of each unit, repetitions coming only after completion of the entire sequence? Another question along these lines (considered in Chapter 7) was: If a task breaks down into natural units, is there a mnemonic advantage to interspersing overt practice at the end of each natural unit of demonstration, as compared with interspersing practice at other points in the total sequence?

Theory

The factor of organization in sequential learning has not been extensively studied in the literature of human learning. The nearest thing to a strictly organizational problem in the classical literature is the "whole-versus-part" problem, in which the question is: If the total task is "forcibly organized" into parts, is it better to master the sequences within parts and string the parts together, or is it better to learn the entire task in sequence? In this classical problem, however, a usual case is one in which no genuine questions of organization are raised. Instead the "parts" are equal-sized subdivisions of the total task, having only nominal distinctions such as "first," "second," "third," etc., rather than genuine organization distinctions. A second area involving organization in the classical literature on human learning is that commonly discussed under the heading of "meaningfulness" of material. This heading is a general "catch-all" category which does not distinguish, for example, between problems of organization *per se* and relatively unrelated problems such as the "association value" of nonsense syllables.

Such nominal units as those found in treatments of the "whole-versus-part" problem should be distinguished from what may be called "natural units." Natural units are present if separate parts of the sequence are differentiated along dimensions inherent in the sequence itself. If such natural parts exist, they give rise to an inherent organization in the learning task in the best sense of the meaning of "organization," which refers to a total structure (or "whole") made up of inherently distinctive parts (or "organs").

In the present theoretical framework, the presence of inherent organization in a sequential task makes the serial learning of its parts easier to accomplish. This expected greater ease of serial learning involves two different propositions about the advantages of inherent organization:

1. A sequential learning task which breaks down into distinctive sub-units simplifies the learning of the total sequence by reducing the task to an easily learned sequence of easily learned sub-sequences.

2. An organized sequence provides a better static perceptual pattern of the task as a whole, the parts of this static pattern suggesting the separate sequential sub-units.

By themselves these two propositions are not very specific, and some expansion on their meaning is desirable.

In the first proposition one implication is that if a total task is subdivisible into perceptually distinctive parts, these parts can be serially learned more readily than the entire sequence simply because fewer items must be sequentially learned. For example, it is easier to learn the six major subdivisions of the central nervous system (cerebral, cranial, cervical, thoracic, lumbar, and sacral) than it is to learn all the separate levels from the most cephalic to the most caudal. Another implication is that the elements within a part are more readily learned than the total sequence, again simply because there are fewer elements per sequence—e.g., the separate parts of the sacral portion of the nervous system are easier to learn than the total cephalo-caudal sequence. A final implication of the first proposition is that the separate subdivisions are sufficiently distinctive as contexts, so that the elements of each subdivision become connected to the distinctive context cues of each subdivision and are not likely to be suggested (i.e., not likely to interfere) in any other context.

In the second proposition above, the main implication is that if the learner can form a single static perception of the total organization, he is in a better position to recall all the subdivisions needed to provide the context cues for the elements within subdivisions. Such static perceptions enable the learner to perceive all the subdivisions as a single entity, or at least to maintain a single perceptual entity that can be maintained while separate subdivisions are attended to. These have been referred to as "perceptual blueprints." The anatomical example provides a good illustration of such "blueprints" because of the well-established perceptual image of an organism as starting with a head and ending with a tail. The six-step breakdown from cerebral to sacral is easily reinstated from this image in terms of the organisms having: (1) brains (cerebral) at the top of the head, (2) the rest of the head (cranial), (3) a neck (cervical) which connects the head to the (4) chest (thoracic), and a (5) loin (lumbar) which connects the chest to the hind-quarters where (6) excretory and genital organs (sacral) are to be found. The single picture of an animal, established in childhood, is a static, "redintegrative," perceptual response which provides a convenient set of distinctive cues to which the subdivision responses can be connected and from which they can be reinstated.

Obviously what holds above for the organization of the central nervous system would hold for a well-structured mechanical device. Thus the drive assembly of an automobile could be broken down into motor, clutch, transmission, and differential, maintaining the same sort of longitudinal organization found in the cephalo-caudal breakdowns of neural anatomy. The main point in either case is that the sequential memorization of order of the total elements is enhanced (a) by the presence of natural subdivisions of the elements and (b) by the existence of an overall perception of the organization of subdivisions.

The same general principles would also hold for the learning of sequences which are not physically as integrated as animals and mechanical assemblies. A poem, for example, might have a mnemonic advantage if it tells a simple story with a plot which breaks down into a few sequential episodes, each of which is related in a separate stanza. Similarly, the sequential details of a new play might be remembered partly by reinstating "perceptual blueprints" based on the stage settings in "Scene 1," "Scene 2," etc. In the latter case, a succession of "perceptual blueprints" is recalled in a sequence involving a small number of separate responses, and each part of the sequence of visual images (i.e., each scene) helps reinstate the characters that were present, where they were located on stage, what action took place, and what was said.

Finally, it should be noted that the method of demonstration can enhance any inherent organization or impose an organization in which very little natural organization is present. A poor demonstration of a complex sequence would merely exhibit the steps in order, whereas a good demonstration would call attention to the general structure, the major units, the sub-units within units, and so forth. The principles involved are much the same as those that apply to the difference between a poorly organized lecture and a well-organized lecture. The contents of a well-organized lecture will be more easily memorized by the lecturer and better remembered by the audience, because numerous devices have been used to segregate the sequentially presented material into easily remembered headings and sub-headings, with some attention to providing an overall pattern into which the major headings fit.

Methods

The methods used were in line with the purposes and theory just described. As already indicated, the experimental variables break down into demonstration variables and utilization variables, the former referring to the construction of the demonstration and the latter to the manner in which the demonstration is combined with overt practice. For the study of demonstration-construction, a new filmed demonstration was prepared. The new filmed demonstration showed the mechanical assembly of a wastegate motor, which is a servomechanism used as part of the supercharging system in airplane motors (see Chapter 6). The complete assembly of this mechanism was considerably simplified for purposes of the experiment, but still required 64 assembly operations in the demonstration and performance of the learning task.

For the study of utilization variables two different filmed demonstrations were used, one of which was the new demonstration just mentioned, and the other a filmed demonstration of the mechanical assembly of an automobile distributor. This latter film had been used in previous studies of utilization factors.[2]

[2]The distributor assembly film was one adapted by Boston University for a previous Air Force Contract, AF 18(600)-1210, from a film made at Yale University under an earlier Air Force Contract: see Part 1 of Chapter 6.

Demonstration Variable

The demonstration variable studied was a procedure designed to enhance the "perceptual blueprint" factor referred to in Chapter 2. The original plan was to expose in the film and describe in narration the completed unit at the beginning and end of each separate sub-assembly. The depiction of the unit at the beginning can call attention only to the major features, since the learner has not yet been through the details; such depiction can nevertheless help establish the general structure of the sub-assembly. At the completion of the demonstration of a unit, however, the parts are familiar, and the completed unit can again be exhibited as a whole, with the relationship among parts alluded to in a more detailed and meaningful fashion.

The difficulty with this original plan was that in the waste-gate motor—and perhaps in most mechanical assemblies—most of the parts of any sub-assembly are concealed inside the housing when a unit is completed. It was necessary, therefore, to devise a demonstration technique which would accomplish a similar result of providing a perceptual blueprint, despite the fact that successive parts hid earlier parts and the entire unit was eventually hidden inside the housing.

The method devised, given the name "implosion" (see Chapter 8), consisted of showing in the film a display of the parts followed by a rapid succession of several-frame "stills" of each part added into their appropriate places. In keeping with the well-known "Phi-phenomenon," this caused the illusion that each part had jumped from the display to its appropriate place in the assembly. Each entire sub-assembly was thus "imploded" in about one minute of demonstration, time being allowed for each part to be perceived before it "jumped" and for its location to be perceived immediately afterward. To this end the implosion sequences were narrated in a fast resume of the sub-assembly, each part being named before it "jumped" and the assembly operation (e.g., location in the structure) being described as it "jumped."

This implosion technique could have been used both before and after the detailed narrated demonstration of the sub-assembly, as in the original plan of exposing the completed sub-assembly at the beginning and end of the detailed demonstration of the unit. It was decided, however, that the implosion sequences would be less meaningful before the demonstration of a given sub-assembly, and they were used only after each such demonstration. As a correlative device, however, each sub-assembly was preceded by the organizing device of showing the place in the main housing in which the unit was to be assembled and by exhibiting the main part around which the unit was built. This brief preamble, combined with the implosion sequence following the demonstration of each unit, constituted the only distinction between the experimental film and the control film, which was identical except for these organizing devices.

Utilization Variables

Two utilization variables were studied. One of them involved the factor of whether a given sub-assembly is demonstrated and practiced twice in succession—versus only once—before proceeding to the next sub-assembly. The other variable was concerned with the question of whether overt practice occurs at the end of each sub-assembly or in the midst of sub-assemblies. These two variables have been discussed in some detail in Chapters 6 and 7, respectively.

The first variable, termed the "Repetitive-versus-Consecutive" variable, was investigated on the hypothesis that if a task is subdivided into inherent organizational sub-units, there is a mnemonic advantage to greater mastery of a sub-unit before proceeding to the next sub-unit. This is the "Repetitive" treatment, in which a sub-unit is demonstrated, then practiced, then demonstrated again and practiced again, etc., before the next sub-unit is introduced. The alternative, "Consecutive," treatment involves one demonstration and then one practice of a sub-unit, followed by demonstration and practice of the next sub-unit, etc. In making such a comparison, of course, the total amount of demonstration and practice should be held constant, so for whatever number of repetitions are used in the Repetitive condition, the same number of times through the entire demonstration-and-practice sequence is required for the corresponding Consecutive condition. In the present studies, this number was limited to two repetitions per unit under Repetitive conditions and two times through the total sequence under Consecutive conditions.

The second variable, termed the "Synchronous-versus-Asynchronous" variable, was investigated on the basis of the hypothesis that if a task is subdivided into inherent organizational sub-units, there is a mnemonic advantage to timing interspersed overt practice so that it comes at the completion of the demonstration of each such natural sub-unit. This is the "Synchronous" treatment, and it used the same stopping points for practice as those used in the Repetitive-Consecutive comparison. The alternative Asynchronous treatment used stopping points which provided the same distribution of temporal lengths of demonstration-practice segments but which came at times as different as possible from those associated with the termination of a natural unit. More specifically, in the task for which this comparison was made (the waste-gate motor assembly) there were four natural units, each with its own temporal duration in the demonstration. These same four temporal durations were used in the Asynchronous treatment, but they were rearranged so as to maximize the lack of correlation between the Synchronous and Asynchronous points at which overt practice was introduced (see Figure 1 of Chapter 7). The general effect of this rearrangement was to put the Asynchronous stopping points about in the middle, rather than at the end, of each natural unit of the task. Only the Repetitive method was used in this comparison; that is, no investigation was made of the interaction between the Synchronous-Asynchronous variable and the Repetitive-Consecutive variable.

As already noted, two different filmed demonstrations were used in the study of utilization variables—the waste-gate motor assembly and the automobile-distributor assembly. Only the waste-gate motor assembly was used in the Synchronous-Asynchronous comparison, but both assemblies were used in the Repetitive-Consecutive comparison.

Summary of Experimental Comparisons

The experimental comparisons described above are summarized in Table 1, which shows all of the treatments involved in the present investigations. It should be understood that the table does not represent a factorial design experiment with some of the cells missing. Rather it merely exhibits in a convenient form the various comparisons made. The data were accumulated in a succession of three different experiments, each of which involved randomization of its subject population over the groups involved, but between which comparisons are less valid because of the lack of such randomization across experiments.

TABLE 1

Summary of Experimental Comparisons

Task, Experiment Number, and Treatment Designation		Organization Variable; Treatment: Repetitive vs. Consecutive Sequencing of Demonstration-Practice Units (Chapter 6)	Synchrony of instruction segments with Natural Units of task (Chapter 7)	Implosion Summary Sequences (Chapter 8)
Automobile Distributor		Ch. 6, Part 1		
Exp. I	(a)	Repetitive	(Synchron.)	(No)
	(b)	Consecutive	(Synchron.)	(No)
Waste-Gate Motor		Ch. 6, Part 2		Ch. 8, Exp. II-B
Exp. II	(a)	Repetitive	(Synchron.)	No
	(b)	Consecutive	(Synchron.)	No
	(c)	Repetitive	(Synchron.)	Yes
	(d)	Consecutive	(Synchron.)	Yes
				Ch. 8, Exp. II-A
			Ch. 7	
Exp. III	(a)	Repetitive	Asynchron.	No
	(b)	Repetitive	Synchron.	No
	(c)	Repetitive	Synchron.	Yes
				Ch. 8, Exp. III

General Procedures Used

The detailed procedures used will not be described in this general report. They have already been described in separate reports dealing with the various experimental comparisons involved in the above design of the investigation. For present purposes it is important only to call attention to general features of the procedures used. These are outlined below.

1. All subjects were run individually. That is, each subject viewed the demonstration and practiced performance alone except for the presence of the experimenter.

2. All subjects viewed each segment of demonstration twice and practiced it twice, followed by a final test in which they attempted to perform the entire task without any demonstration.

3. Performance in each practice and in the final test was scored in terms of both number of correct responses and time to complete such responses as were attempted. In all cases a maximum time limit was imposed, such that if the subject had failed to finish, his attempts were terminated. The time limits were set so that few subjects had to stop before they had reached their limits of achievement.

4. In practices on sub-units, the subject always started with an assembly which was correctly put together up to the point of his last attempt, regardless of how many errors of omission or commission he might have made previously. This prevented the compounding of errors across units. In the final test, however, the subject was "on his own" throughout and could compound errors where possible.

5. Dependence of performance on sheer manual dexterity was kept at a minimum. Thus if a particular assembly operation consisted of selecting the right machine screw and screwing it into the right hole, S was scored not only in terms of whether he made the appropriate attempt (correct response), but also how long it took him to make the correct attempt (time score). However, the time required to get the screw tightened by means of the manipulation of the screwdriver was eliminated from the time scores. For the same reasons, the physical mechanisms used for practice and testing had been modified to eliminate manipulative problems where parts normally have very fine tolerance. That is, all tight fits were somewhat reduced (e.g., a shaft would be reduced slightly in diameter) such that fairly naive subjects could fit parts together if they understood where the parts belonged in the assembly, regardless of manipulative skills.

Results

The detailed findings of the present studies will not be reported here, since the present purpose is to abstract the main findings as a basis for the interpretation section. The main findings can be briefly abstracted as follows:

1. The Repetitive method of utilization was superior to the Consecutive method at all stages of practice and in the final test in the case of the automobile-distributor assembly task.

2. The Repetitive method of utilization was superior to the Consecutive method only during practice in the case of the waste-gate motor task, the two conditions being essentially equivalent in the final test.

3. The Repetitive method is characteristically associated with a decline in performance from second practice to final test, whereas the Consecutive method is characteristically associated with an improvement in performance from second practice to final test. This does not imply that the latter method is necessarily superior in the final test; it only means that, in general, the Consecutive method tends to show a continued improvement from practice to test, whereas the Repetitive method tends to show a decline from practice to test despite its general superiority during practice. In no case, however, was the Consecutive method superior in the final test. It improved, but ended up either inferior (automobile distributor) or about equal (waste-gate motor) in the final test.

4. With the task used (waste-gate motor) the Synchronous method of practice was not found to be particularly superior to the Asynchronous method.

5. The "Implosion" device is a filmed-demonstration technique of considerable mnemonic advantage in teaching mechanical assemblies, even when the demonstration otherwise has attempted to maximize all of the pictorial and narrational methods which enhance the ease of learning of a well-organized task. This result should be considered in the light of the fact that a poorly organized demonstration was not used. Thus, if an inferior demonstration had been prepared, the effects of adding the Implosion sequences would be expected to be even larger.

6. A well-organized sequential task, when presented in a well-organized demonstration, can be learned to a fairly high degree of proficiency with only two overt practices, even though it is relatively long and complex (i.e., 64 assembly operations in the waste-gate motor assembly). This conclusion is based on the fact that the waste-gate motor assembly was surprisingly well learned under all conditions, despite its large number of sequential items. The average performance on the final test was in the neighborhood of about 90 per cent correct responding in assembly, without any prompts from the filmed demonstration.

The foregoing conclusions represent the main findings of the present studies. They will be alluded to in the interpretation which follows, as will certain subsidiary findings not mentioned above. Such subsidiary findings are covered in previous reports, and will be brought in where relevant in attempting to interpret the overall results of the present research.

Interpretation

One of the findings which requires interpretation is that, whereas the Repetitive method was superior to the Consecutive method in both practice and test with the distributor assembly, it was superior only during practice with the waste-gate motor assembly. In the latter task the significant superiority of the Repetitive group during the second practice period disappeared in the final test. This change in relative performance appeared to be due not so much to the Consecutive group's catching up with the Repetitive group as to the Repetitive group's failure to improve between second practice and final test.[3] This failure to improve was characteristic of all of the Repetitive groups in the waste-gate motor task; that is, it held whether Implosion was present or absent and it held for Asynchronous as well as Synchronous groups. On the other hand—see Table 3, Chapter 8—both Consecutive groups (i.e., with and without Implosion) showed improvement from second practice to final test.

One possible explanation for this discrepant result between the two kinds of tasks (automobile distributor versus waste-gate motor) is that, whereas only two repetitions were used for each task, assembling the waste-gate motor was a more complex task involving over twice as many operations. Thus two repetitions may not have been enough for the advantages of the Repetitive method to take hold sufficiently to show up when the final test conditions were introduced. The shift to final test conditions is a transfer-of-training problem for both the Repetitive and the Consecutive methods. This transfer problem is greater for the Repetitive groups, however, because they had only one practice at moving from one unit to the next, whereas the Consecutive groups had two such practices, while at the same time having no practice at going back over the same material immediately after practice. To this extent the Consecutive method is more similar to the final test procedure, and less transfer decrement would be expected. Thus it could be that with limited experience the organizational advantages of mastering a distinctive unit before proceeding to the next would be offset by the transfer disadvantages of having less opportunity to practice the cueing of the next unit to the completion of the preceding one. With more extensive practice the intra-unit organization factor might be relatively more effective. By the same token, if practice is held constant and the task is varied in complexity, it would be expected that the simpler task would show more of the advantages of the Repetitive method than the more complex task, which would account for the fact that the shorter distributor task showed more effect of the Repetitive method than did the waste-gate motor task.

The chief argument against this interpretation is that the waste-gate motor task was not poorly learned, despite having 64 assembly operations and only two demonstrations and practices. The correct-response level was close to 90 per cent by the second practice, even for the Consecutive method. This tends to argue against an insufficient-practice interpretation, although it does not rule it out. In any case, it must be recognized that the Repetitive method will always face a larger transfer-decrement problem than the Consecutive method.

Other explanations of the disparity in results between the two different assembly tasks are also possible. Thus, it may be that if a task is inherently well organized, any attempts to increase organization factors further will add very little to what the

[3] Actually the Repetitive group dropped slightly in performance rate, although the drop was not statistically significant.

subject does, regardless of methods of combining demonstration and practice. Against this explanation is the fact that the equivalence of the two methods in the final test comes about, not from an equivalence throughout practice and testings, but rather from an apparent falling off of the Repetitive group's superiority when test conditions were introduced. Thus it appears that the Repetitive groups lost something when the test was given. Another possible explanation is that the Repetitive method makes performance too easy, so that the motivation to learn on the second time through the demonstration stimulates less effort by the learner. The Consecutive subject is always moving to new material, whereas the Repetitive subject goes back over immediately familiar material and may feel that "he knows it cold," and thus does not need to attend to the second demonstration or rehearse implicitly during the second practice.

It is, of course, impossible to choose confidently among the various possible explanations without further research. As a best guess, however, the following interpretation is proposed here. In the first place, it should be recalled that the waste-gate motor task was carefully selected for its inherent organizability, whereas this was not the case in the distributor task. Also, the filmed demonstration (and its narration) was prepared with an attempt to emphasize the inherent organization of the task. Given these conditions to begin with, the task would be expected to be fairly readily learned by any combination of demonstration and practice. A procedure which minimizes intra-serial interference by maximizing organization within sub-units will sacrifice to some extent the organization of the task as a whole. The learner would therefore be expected to suffer a temporary setback when conditions of performance are shifted to those of performing the task as a whole. By contrast, a training procedure which emphasized the organization of the task as a whole at the expense of organization within sub-units would show somewhat poorer performance during training but little setback when performance was shifted to the test conditions of executing the task as a whole. All differences would be relatively small, however, because of the inherent organization. This would account for the results obtained with the "Repetitive-Consecutive" comparison in the waste-gate motor task.

If, on the other hand, the inherent organization into sub-units was weak, and an organization had to be imposed to some extent, any setback due to shifting from training conditions to testing performance on the whole task could be more than offset by the minimization of intra-serial interference. This would account for the fact that the Repetitive condition was superior in both practice and final test in the distributor-assembly task. The conclusion then would be that, with inherently well-organized tasks, the Consecutive method will be at least as effective as the Repetitive method, but with less well-organized tasks the Repetitive method will be superior.

This interpretation would also fit in with the results of the "Synchronous-Asynchronous" comparison with the waste-gate motor. Here again the inherent organization of the task would make integration within sub-units fairly easy, even with Asynchronous practice; whereas the use of Asynchronous practice would tend to facilitate mastery of the organization as a whole, due to practicing terminal portions of one sub-unit in the same context as beginning portions of the next sub-unit. The equivalence of results at all points of training and in the final test for this comparison would then be interpreted as a cancellation of effects—gains from within-unit organization (Synchronous method) being exactly offset by gains from between-unit organization (Asynchronous method).

Alternative interpretations of the essential equivalence of the Synchronous and Asynchronous procedures are possible, however. One such interpretation is that the Asynchronous procedure had the advantage of breaking the longer sub-units into two parts, allowing practice to be interspersed at times likely to be more strategic from the standpoint of a "Demonstration-Assimilation Span" unit (see Chapters 2 and 5). This interpretation would, however, have to make some additional assumptions about the relative effectiveness of the different factors involved.

It could be argued, for example, that overt practice helps to "consolidate" perceptual learning in the middle of a complex sequence, where intra-serial interference is maximum. Thus the subject might learn the beginning and end of a lengthy sub-assembly entirely on the basis of the demonstration, but might be confused in his perceptual recollections of the middle portion. Stopping for overt practice at such points might tend to eliminate confusion by permitting the subject to practice overtly while he can still remember most of the steps up to that point. The best support for this sort of alternative interpretation in the present study comes from the fact, revealed by more detailed analysis, that the data for the Asynchronous method show a greater number of correct responses than for the Synchronous method in the two longest units of the waste-gate motor task (the first and the third). If more practice had been allowed, on the other hand, the organizational factors present in these longer units would be expected to more than offset any temporary advantage of stopping to practice in the middle. Thus the results might have been quite different if three or four demonstration-and-practice sequences had been used instead of only two. Again the situation is one in which further research is required before a definite conclusion can be drawn.

The results are less ambiguous with regard to organizational factors built into the demonstration. Thus the utilization factors studied are subject to the possible counteracting effects of factors which aid learning in one respect but interfere with learning in another respect. When organizational advantages are incorporated in the demonstration itself, however, the utilization procedures can be held constant and the organizational factors varied. In the present instance the use of the implosion device proved to be effective as a means of enhancing within-unit organization. It accomplished this despite the fact that the task itself was an inherently well-organized one and despite the fact that the demonstration—even without implosion—was designed to maximize organizational factors.

One of the most surprising results of the present studies was the fact that acquisition progressed so rapidly, despite the complexity of the waste-gate motor task. The research had been undertaken for the purpose of studying the effects of organizational factors in minimizing the intra-serial interference associated with lengthy serial learning tasks. For this purpose it was deemed essential that a lengthy task be selected, and the waste-gate motor was adopted with some misgivings about its length (64 assembly operations), which was considered possibly excessive for purposes of experimentation. Thus the serial learning of a nonsense-syllable list of 64 items would be expected to require a great many repetitions, whereas the present experimentation was to be based on only two demonstration-and-practice sequences. It was assumed from previous experience with mechanical assemblies, of course, that the serial learning would proceed much more rapidly than in learning a series of nonsense syllables. But it was not expected that the level of correct responding would be in the region of 90 per cent by the second practice. It was partly for this reason that the filmed demonstration was prepared with a view to using every possible means of enhancing inherent organization and minimizing intra-serial interference.

Thus the general effectiveness of the filmed demonstration can be regarded as one of the products of the research. No control was used, of course, by which a different and less-effective demonstration of the task could be compared with the demonstration actually prepared. But the general effectiveness of the teaching procedure that was used, despite the difficulty of the task, is perhaps of some significance—not only because it may have been a factor attenuating the effects of the experimental variables studied, but also because it indicates that complex assemblies can be taught without a large number of demonstration-and-practice sequences and without any individual guidance by an instructor.

From this latter standpoint it may be of some interest to list the principles followed in the preparation of the filmed demonstration of the waste-gate motor assembly.[4] Again there is no control on which principles are useful and which are not, but at least the list comprises the policies followed in the preparation of what apparently was a successful filmed demonstration of a mechanical assembly task. The principles are labeled according to whether they are pictorial, narrational, or both.

1. (Pictorial). The assembly demonstration should be taken over the shoulder of the demonstrator to maximize the similarity between demonstrational perceptions and practice perceptions and to minimize any distracting effects of attending to the demonstrator. The face of the demonstrator should never be seen (cf. Roshal, 1949a,b: Part II of this volume).

2. (Pictorial). The entire set of parts should be displayed at the outset and left in the same positions until they are picked up and relocated in the assembly itself. This procedure gives an opportunity for seeing all the parts at once, for discriminating among the different parts, for getting a perception of the relative sizes of different parts, and for providing a unitary perceptual memory of all of the parts, which may be helpful in subsequent attempts to select the right part by "scanning" this memory image.

3. (Pictorial and narrational). As each part is selected for assembly it should be displayed by itself while it is being named and described. This should be managed

[4]It is of interest to compare this list of principles with those suggested for demonstration films (though not then implemented in a specific film) by May (1946, 1958); see concluding chapter of this volume.

by (a) showing the selection from the display, (b) placing the part next to the partially completed assembly, (c) then viewing in close-up the part while its distinguishing features are described (cf. Wulff and Kraeling, Part II of this volume). This method insures that the part is seen in relation to the entire display, in relation to the assembly, and in detail while it is named and described before the assembly operation is demonstrated.

4. (Pictorial). At each stage of assembly, the structure assembled up to that point should be kept in view as much as possible. This not only helps to give a "baseline" for relative sizes of parts, but also maintains continuous rehearsal of the visual perception of the structure being assembled. From the standpoint of establishing a "perceptual blueprint" of the structure, this procedure is of considerable importance.

5. (Pictorial). The arrangement of a display and the placement of a part for individual inspection should maximize the perception of the total form of each individual part. This is a rather subtle principle to apply, but can be illustrated by the case in which a thin disk could be displayed (a) from a side view, in which case it would look like a line, (b) from a top view, in which case it would look like a circle of unknown depth since it could be either a disk or the top of a cylinder, or (c) from an angular view, in which its shape would be elliptical but the normal perception would be that of a thin circle tipped at an angle. The point is that the actual structure of an object is a perceptual inference from the visual stimulus presented, and the pictorial principle involved is that there is generally a view which maximizes the amount of information about total structure conveyed by a still shot of an object. This is the view which should be used both in the original display and in the close-up inspection mentioned in (3) above.

6. (Pictorial). A corollary of (5) is that there is often an advantage to displaying a given complex part from several angles. Thus, if the nature of the back of an object cannot be inferred from a view of the front, both sides must be shown in order to establish the properties of the object as a whole. In such cases the demonstrator should turn the object around so that it can be viewed from various angles. In this way a total perception of the object is established, such that its appearance from one angle suggests, through cross-conditioning of perceptual responses (see Chapter 2), its appearance from another angle.

7. (Pictorial). A close-up of any part should never be used without an initial and final longer shot which establishes the relative size of the unit compared with the rest of the assembly.

8. (Narrational). Each part should be given a name (as contrasted, for example, with calling it "this part") which as nearly as possible describes either some characteristic of the part or some characteristic of its location in the assembly. This principle frequently cannot be applied where it is deemed important to teach technical names. In the present study some technical names were dropped in favor of using more meaningful labels. (For example: the technical expression, "pinion gear" was dropped in favor of the more meaningful expression, "smaller gear.")

9. (Pictorial and narrational). When parts are similar in appearance but go in different locations in the assembly, they should be placed together visually and their differences described verbally. Thus, in the present demonstration there were machine screws of different sizes and with different head shapes. At the appropriate point in the demonstration, all of the different types of screws were placed side by side and attention called to their differences.

10. (Pictorial and narrational). In general the narration should always slightly precede, in its description, the visual portrayal of an assembly operation. This rule applies only where movement is involved and is probably unimportant where narration describes a stationary object. But when a part is being placed in its proper position, the subject should be set verbally for what he is to observe visually. Otherwise he may attend to some entirely trivial aspect of the visual movement and miss completely the significant part of the assembly operation.

11. (Narrational). Any feature of a part which can serve as a reminder of the nature of an assembly operation or the location of the part in the total structure should be named and described. For example, in the waste-gate motor, one of the wiper arms of the potentiometer was longer than the other, and this arm was the one which had to make contact with the resistance winding. In such a case it would be a very inferior procedure to say: "The wiper goes on like this," but a highly effective procedure to call attention to the difference in length and note that the longer arm is the one that should be in contact with the winding.

12. (Narrational). A context cue should be emphasized for each possible sub-assembly or sub-sub-assembly into which the total assembly can be organized. This context cue should be some feature of the sub-assembly which distinguishes it from all others and which remains more or less in evidence throughout the assembly of the sub-unit. This feature should be named and identified as the major part in memorizing the particular sub-structure. Thus, in the waste-gate motor, the solenoid brake assembly was described in terms of the "solenoid," the "solenoid shell," the "solenoid cover," and so forth. Emphasis on such cues is in line with the basic hypothesis of the organizational theory used here; namely, that intra-serial interference is eliminated to the extent that (a) cues in different parts of a sequence can be made distinctive, and (b) the appropriate responses can be connected only to their own distinctive cues.

PART II

MOTOR, PERCEPTUAL, AND ACADEMIC SUBJECT-MATTER LEARNING

INTRODUCTION TO PART II

The studies reported in this part dealt with a variety of learning tasks. All of them employed some form of presentation of instruction by film, but the role of the film instruction in relation to subsequent student performance differed considerably among the studies.

In the study reported by Wulff and Kraeling in Chapter 10, the main focus of attention was on the guiding of implicit responses; overt performance by students was not required until the test situation. This study, like most of those in Part I, dealt with the serial manual task of assembling a mechanism (an ignition distributor) and employed the same film materials that were later used for the studies reported in Chapter 3 and the first part of Chapter 6. The theoretical orientation of this study was in terms of cue factors that foster the acquisition and effective utilization of implicit or perceptual discriminative responses by the learner through prior familiarization with the distinctive cues of the parts to be assembled. The most interesting aspect of the experimental data is the comparison of procedures which gave prior familiarization with the parts before any training on how to assemble them, versus a condition in which familiarization with each part was given a step at a time just prior to its use in the assembly.

Chapter 11 reports a very interesting study by Roshal, which used a different kind of sequential manual task, knot-tying. Roshal was concerned also with the perceptual aspects of the task and, in particular, with variables which would enhance the similarity between the cues provided to the learner during the demonstration and those that would be present later, during attempted performance of the task. Some of the experimental treatments introduced opportunity for overt response (student participation) during the filmed instruction, and several other characteristics of the visual character of the demonstration were also manipulated. The latter represented variables that were hypothesized to affect the degree of transfer from implicit (or overt) responding during the demonstration to later overt performance involving visual cues with varying degrees of similarity to those present in the several experimental versions of the demonstration. The variations included the so-called "subjective" versus "objective" camera angle, moving-picture depiction of the narrated demonstration versus depiction by a narrated series of still pictures, and delineation of the task sequence both with and without inclusion of the demonstrator's hands.

In the studies reported thus far, the introduction of overt practice has been conceived as complementing the perceptual-symbolic responses learned during a demonstration, and the tasks used demanded a serial ordering of responses that would not be likely to occur without such prior demonstration. By contrast, the studies reported by McGuire in Chapters 12, 13, and 14 were concerned with a "cyclical" task, pursuit-rotor tracking. Skill in this perceptual-motor task is ordinarily acquired by practice which can be carried on without any necessary preceding instruction other than that needed to define the task objective of "keeping on target." The task, extensively studied in the psychological laboratory (e.g., Ammons, 1947b, 1955), appears to be one of almost sheer eye-hand coordination, involving no particular memory factor, or at least imposing no requirement for the serial learning and integration of a sequence of dissimilar component acts. The studies reported here introduced the innovation of using filmed instruction to guide the learner's subsequent practice on such a task, and attempted not merely to improve performance but to do so by teaching intervening postural responses and habits believed to be conducive to efficient task performance. The roles of demonstration and overt practice could in a sense, however, be considered to be reversed, as compared with some of the preceding studies; that is, the perspective was of using filmed demonstration to aid in the acquisition of a skill through practice, rather than introducing overt practice to augment learning from a demonstration.

In the study reported in Chapter 12, the experimental treatments involved merely a comparison of skill acquisition and postural responses during practice for groups who practiced with or without preceding film demonstration. The significant aspect of the study lies not so much in the overall comparison, however, as in the analysis of performance in relation to several hypotheses concerning specific factors, some of which, rather than making positive contributions to learning, were hypothesized to produce detrimental effects on practice performance as a result of using the film demonstrations. These factors included leveling effects, massing effects, and interference attributed to anxiety arousal.

The further study reported by McGuire in Chapter 13 represents a more extensive experimental inquiry that employed a rather ingenious multi-factor design to investigate four principal variables. Two of these, analyzed in Section I of the chapter, were (1) the number of repetitions of the demonstration and (2) the inclusion of accompanying descriptive narration versus silent demonstration of the task. The further variables dealt with in Section II of the chapter were (3) slow motion versus standard-speed depiction of the demonstrations, and (4) massed versus spaced showing of the demonstrational films, as well as the interactions of these variables with factor (2), above.

The study of slow motion reported in this Section II is of particular interest because, while somewhat limited in scope, it thus far stands virtually alone in the experimental literature as an inquiry in which this potentially important demonstrational-cue variable has been manipulated as a perceptual mediator of skill acquisition. The paucity of studies dealing with slow motion as a cinematic-perceptual factor in film instruction is quite surprising in view of the fact that slow motion is not only relatively easy to implement cleanly as an experimental variable, but is also one of the very few attributes (if, indeed, not the sole attribute) that is genuinely unique as an instructional potentiality of motion pictures.[1]

The study reported by McGuire in Chapter 14 is based on separate item-by-item scoring of the responses of 16 control and 16 experimental subjects to each of the eleven separate instructional points illustrated in a film demonstration (which was similar to that used in the Chapter 13 study). This analysis is concerned with differences in the effects of these points as a function of their serial position in the demonstration and, in particular, their temporal proximity to several "reward" sequences that were interpolated following three major subdivisions of the total demonstration. This study is of interest, especially in view of the general paucity of studies—noted by Cook (1960a) and by Miller et al. (1957)—that have explicitly introduced manipulation of reinforcement or reward variables in experiments on instructional media. (Also see the unpublished report by Kimble (1955a), abstracted in the Appendix.)

The two studies reported by Kimble and Wulff in Chapters 15 and 16 are both concerned with variables related to the kind and amount of overt student-response exercises that might maximize the learning of a symbolic-perceptual task—reading the kind of graduated (logarithmic) scales used in the slide rule. The primary variable investigated in Chapter 15 was "response guidance"—that is, the use of hints, cues, or prompts that could be given to the learner to increase the likelihood of his practicing correct rather than incorrect responses during each of a number of practice exercises interspersed between short segments of film instruction on the slide rule. In addition to the overall beneficial effect found for using such "crutches" or "prompts" to guide the learner to the correct response, the analyses presented compare the effect of this procedure on harder versus less difficult material, and on positive transfer to new examples not previously encountered (as well as on the scale-settings used during practice). An analysis was also made of the kinds of errors that were affected by the guidance procedure to ascertain whether, as intended, its success was due primarily to reduction in errors of the kind which the guidance cues, employed in the film, had been expressly designed to help the student avoid.

The further investigation reported in Chapter 16 posed, for this kind of instructional task, essentially the question investigated by Gates (1917) in his early studies with verbal-learning material—namely, the question of optimum proportions of demonstration and practice. The proportion of practice exercises (as contrasted to demonstration examples) included three categories (zero per cent, 50 per cent,

[1] Another of the few attempted studies of slow motion known to the Editor dealt with an acrobatic maneuver (the "Kip"). This study was initiated by F. D. Sheffield around 1948 in connection with the Yale University Motion Picture Research Project but, following the execution of initial slow-motion and parallel normal-speed cinematography, had to be suspended for lack of continued funding of the project. Recently a further investigation using a similar learning task has been initiated by Rubino (1961) at the University of California, Los Angeles.

and 100 per cent) in a first experiment, and five categories (ranging from zero to 100 per cent) in a second experiment. The results of the two experiments taken in combination are not fully conclusive but, at a minimum, the trend of the data is compatible with the assumption that some intermediate proportion of time devoted to student practice is generally superior to either a very large or a very small proportion.

The experimentation reported by Lumsdaine, Sulzer, and Kopstein in Chapter 17—like that reported in Chapter 10 and much of the analysis of factors considered in Chapter 11—do not deal explicitly with overt responses of the student (except as these are observed in testing to measure the effectiveness of instruction). Its primary relevance is to the cueing of implicit responses through manipulation of attention-focusing cues produced by animation techniques (in Section I of the chapter), and on the effects of repetition of instructional examples (analyzed in Section II of the chapter). The animation techniques involved attention-directing cues provided by animated arrows and labels that were used concurrently with the main expository portions of the film (which gave instruction on how to read micrometer settings). It is interesting to compare the contributions made by these attention-directing devices with the results reported by Wulff and Kraeling in Chapter 10 for the somewhat similarly-functioning cues used "to familiarize" trainees with the salient features of component parts of a mechanism (see above). In addition, both this use of animation (to provide relevant cues for fostering correct implicit responses in reading micrometer settings) and also the related use of animation cues for visual labeling of parts in teaching nomenclature, are of interest in relation to other studies reported in this volume.

The former use of animation for guiding implicit responses would appear to bear considerable functional similarity to the "guidance" technique described by Kimble and Wulff in Chapter 15 for fostering correctness of the overt responses made by students at frequent intervals between segments of an instructional-film presentation. It is also of interest in relation to McGuire's study of additional verbal information furnished by narration added to a silent film (Chapter 13). The animation-labeling techniques studied here (and also the just-noted aspect of McGuire's study) are further of interest in relation to some of the data on verbal versus pictorial stimulus-term representation reported by Kopstein and Roshal in Chapter 20. Similarly, the study of diminishing returns with added repetition, presented by Lumsdaine, Sulzer, and Kopstein in Section II of Chapter 17, may help to clarify in part the negative findings on added demonstration reported by McGuire in Chapter 13. It is also of interest—particularly in terms of the interactions found between repetition, learner intelligence, and difficulty of the material—in relation to other studies (e.g., McTavish, 1949) concerned with repetition as a factor in the effectiveness of instructional media.

The final two chapters of Part II report investigations conducted at Boston University to determine the influence of several factors on the effectiveness of student response in the learning of "academic" subject matter or, in terms of current parlance in the auto-instructional field, "connected-discourse" material. The 1951 study reported by Michael & Maccoby in Chapter 18 was conceived as a follow-on to the World War II experiment by Hovland, Lumsdaine, and Sheffield (1949). The experimental variations introduced by Hovland, Lumsdaine, and Sheffield left undetermined the relative amount of contribution, in the student-response procedures they studied, ascribable to sheer practice effects and to motivating effects of being required to respond. An analysis of their data in relation to the effect of motivating instructions suggested that both factors were operative, since student participation had less effect when students were given more extrinsic motivation (by being told that they would be tested after the film instruction) than when they were not so motivated.

Michael and Maccoby, in the study reported in Chapter 18, made a further analysis of the relative contribution of the two factors of "motivation" and "practice" by giving practice on only half the material covered by a film. From the results of the experiment reported in this chapter, they concluded that virtually all of the effect of the student-response exercises was represented by straight practice effect, since they found no transfer to the portion of the subject matter that was not specifically practiced. (A similar finding was obtained in a later replication with paired-associates material by Kanner and Sulzer: cf. Chapter 28 in Part III). However, in the further study reported by Maccoby, Michael, and Levine in Chapter 19, results obtained with a less interesting subject matter led to a modification of this conclusion. In this case, no appreciable effects attributable to the motivation factor were found when subjects were told that they would be tested, but significant motivational effects were found when this extrinsic test motivation was not introduced. In Chapter 19, several other experimental analyses were also made to amplify the previous experimental results, partly in reference to selective-attention effects similar to those earlier studied by May and Lumsdaine (1958, Chapter 7).

A second major factor investigated both in the study reported in Chapter 18 and in that reported in Chapter 19 was "feedback" or "knowledge of results"—perhaps more accurately described as "knowledge of the correct response." However named, the dependence of gains due to practice exercises upon such feedback bears an interesting relation to further studies (e.g., those reported in Part III) on the importance of the factor called "confirmation" as compared with the importance of "prompting" to insure correct responses. A third factor, to which a very considerable amount of subsequent research has been devoted (particularly with auto-instructional programs) was also investigated in the earlier (Chapter 18) Michael and Maccoby study —namely, the comparison of so-called "overt" practice, employing, here, written responses by the student, with "covert" responding, in which the student is instructed merely to "think" his response rather than make it overtly.

In relation to both theoretical considerations and empirical results, the reader may note several other interesting parallels between the investigations reported in several chapters of Part II (dealing with manual and perceptual skills as well as substantive academic subject-matter) and similar theoretical considerations and empirical findings reported for paired-associates learning in Part III. These parallels are also in evidence, for various subject matters involving substantive "connected-discourse" learning, in current investigations of teaching-machine programs and other forms of auto-instructional materials.

CHAPTER 10

FAMILIARIZATION PROCEDURES USED AS ADJUNCTS TO ASSEMBLY-TASK TRAINING WITH A DEMONSTRATION FILM[1]

J. Jepson Wulff
Doris Kraeling

While few training tasks can be classified as requiring only one kind of learning, many training tasks are of such a nature that the principal learning problem can be identified. Thus a principal learning problem in aircraft identification is stimulus-discrimination learning. One way to develop a training instrument is to classify the training task in this way, and then to select a training technique on the basis of its applicability to the kind of learning problem that characterizes the task. With this approach, training techniques such as student participation, familiarization, animation, etc., are conceived as means which can be utilized to facilitate different kinds of learning, and such techniques are used only when they can be justified as appropriate for the particular kind of learning problem involved.

In this study an attempt is made to identify a specific kind of learning problem and then to test the utility of a specific training technique for that kind of problem. In order to employ a research approach based on the idea that different kinds of training solutions are required for different kinds of training problems, it is necessary to attempt first to identify in learning terms the type of problem that is characteristic of the task for which training will be attempted. Then it is necessary to construct training conditions that can be expected to facilitate learning when the identified learning problem is present. By training conditions we mean stimulus conditions that are employed during the learning process specifically to shape the behavior desired. In this study the training conditions were designed to have effect

[1]This work was done under Contract No. 18(600)-35 between Yale University and the U.S. Air Force, monitored by Dr. A. A. Lumsdaine. The monitoring responsibility for the contract, initiated under the Audio-Visual Research Division of the Human Factors Operations Research Laboratory, was later transferred to the Training Aids Research Laboratory (AFPTRC). The Yale contract was under the general direction of Professor Mark A. May, Principal Investigator, and the contract work was supervised by Dr. Fred D. Sheffield. This report was prepared by Dr. J. Jepson Wulff. It is a revision of an unpublished Training Aids Research Laboratory Staff Research Memorandum by Wulff, Sheffield, and Kraeling (1954). The authors are indebted to Dr. Kenneth H. Kurtz for assistance in conducting part of the experimental work. Wulff is now with Psychological Research Associates, Arlington, Va., and Kraeling is at the University of Vermont.

prior to any possible effects of reinforcement. That is, the stimulus conditions employed as training conditions were designed to foster correct responses as often as possible on the first learning trial, and thus to shape behavior even before reinforcement effects could be brought to bear. A description of the experimental training task, an analysis of this task for the purpose of identifying a principal learning problem embedded within the task, and the rationale for the training conditions selected for experimental evaluation are discussed below.

In developing the rationale for the learning analysis it is necessary to differentiate several meanings of the word "stimulus." While stimuli become associated with responses in the associative paradigm, this certainly does not mean that stimulus objects become associated with responses. One way to clarify the nature of the stimulus event in associative learning is to propose that it is implicit responses to individual features of stimulus objects which function as stimulus events, and that these implicit responses as stimuli become associated with overt responses. In this paper we will call the stimulus features which elicit such implicit responses "cues," in order to differentiate them from stimulus objects. This kind of precision is useful, of course, only if the relationship between the stimulus object and the implicit responses is a variable one, and that is just what is proposed here. Thus, several experiments (e.g., Eckstrand, 1952; Kurtz, 1953; and Wulff, 1954) have shown that, for different learners or for one learner under different conditions, different cues will be utilized in learning even though the stimulus object remains unchanged. For example, Eckstrand (1952) has shown that when both red and round are unique features of a stimulus object, and when the object is a stimulus item in paired-associates learning, the learner may associate the required response with either red alone or round alone.

For training research the important implication of this notion that the associated stimulus is an implicit response is that a training device should insure not only that trainees make correct overt responses, but also that they make these overt responses to appropriate cues or features of the stimulus object, or more precisely, to reliable implicit responses to distinctive cues. When the stimulus object has many stimulus features, a reliable implicit response (or cue) may be defined as one which will be adequate in the performance situation. Conversely, an inappropriate cue would be one, for example, which discriminates stimulus objects during training, but which is not present or is not discriminative in the performance situation.

Problem

One kind of training task that frequently involves unfamiliar stimulus objects with many stimulus features is the task of learning to assemble mechanical units. Commonly the overall aspect of a new mechanical part is unfamiliar to the learner, although it may have several familiar features such as threaded holes, bearing surfaces, slots, pins, etc.

According to the present stimulus notion, the learner, without guidance, may respond to any or all of the familiar stimulus features of such stimulus objects as cues. However, correct performance of the task may require the learner to associate overt responses with specific features of each stimulus object. Thus a mechanical part may have several predominant features such as a round outline, black color, etc., while assembly of the part may require the association of a response with an obscure threaded hole as a cue. Or, the predominant features of a part may be poor cues to employ because they are the same as the predominent features of another part, giving rise to confusion. This suggests that a training procedure which will facilitate learning to perform the assembly is one which encourages the learner at the outset to associate overt responses with the same cues as those he must finally use in performance of the task, rather than with others which may be prepotent but not useful in actual performance (cf. Chapter 11).

One way to encourage the learner to make responses to appropriate stimulus features is to familiarize him with the mechanical parts prior to instruction in assembling them by pointing to the relevant stimulus features of each part and by requiring him to make overt responses to these as cues by pointing to them.

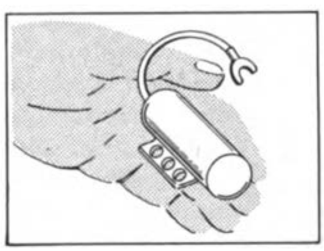
The assembly of a mechanical unit requires at least two major kinds of behavior: the selection of the parts to be assembled at a particular stage in the assembly, and the assembly of the parts after they have been selected. According to the present stimulus notion, the individual trainee identifies a part by the implicit responses which he has made to its features, which thus become cues. Therefore, training that encourages the use of appropriate features as cues should improve subsequent performance in selection of the parts for assembly in the proper order. Likewise, assembly of the part, once it has been selected, should be facilitated if, during training, assembly responses have been associated with appropriate stimulus features. That is, if trainees were given suitable familiarization during assembly training, then it could be expected that they would be likely to attend to useful cues from the outset of training, and that the effectiveness of a training trial with familiarization would be greater than the effectiveness of a training trial without familiarization. Trainees given no familiarization would be expected, during assembly training, to attend to obvious cues which would not always be appropriate, and thus their selection and assembly performance would be inferior to that of trainees given familiarization-training.

The present experiment was performed in order to test the utility of a familiarization procedure as a training condition for a task that included as a principal learning problem the problem of utilizing appropriate discriminative cues.

Method

Training Materials

The assembly of an automobile-ignition distributor was the learning task employed in this study. Two demonstration sequences were prepared on 16-mm. black and white movie film to teach this assembly. One film version, the basic version, was 22 minutes in length. In this version the entire assembly was shown step by step in sequence by means of medium close-up shots, and the action was explained by a magnetically recorded narration. As needed for the assembly, the mechanical parts were introduced one at a time by means of close-up shots. The other film, the familiarization version, was about 30 minutes in length. In this version the medium close-up shots showing assembly steps were identical to those in the basic version. The close-up shots of individual parts were different, however. In the familiarization version these close-ups included action and narration which directed the viewer's attention to those features of each part to which it would be necessary to attend during assembly. Thus, if a part had a threaded hole and a square hole which had to be used to attach subsequent pieces, these were pointed out in the close-up shots. However, during these close-ups reference was not made to features which are not essential to note in performing the assembly. Sketches representing sample close-ups from each version are compared in Figure 1. The part shown was called the "condenser." For the basic version this part was shown without pointing and without comment about its characteristics; for the familiarization version a pencil was used to point to the features indicated by pencil in the figure, and these features were described in the narration as they were pointed out.

PICTURE	NARRATION	ACTION
 BASIC VERSION	...we are ready to attach the condenser.	None

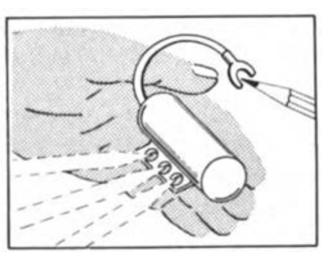
PICTURE	NARRATION	ACTION
 FAMILIARIZATION VERSION	... we are ready to attach the condenser. Notice the three small holes in its supporting base and the slot in the flat terminal at the end of its wire.	(A pencil is used to point to the features referred to in the narration, and shown at the left.)

Figure 1. A sample No-Familiarization close-up from the basic version of the demonstration film compared with the analogous close-up from the Familiarization version of the film

In addition to these filmed sequences, training materials for familiarization without film were prepared by mounting each of the actual parts of the distributor separately on masonite boards about four inches square. These boards were numbered and arranged in a rack so that any one part could be easily located by the experimenter.

Design of Training Procedures

Three different training procedures were devised utilizing these training materials. Assembly training was essentially constant for all three of these procedures; for all procedures it was accomplished by showing the film sequences that were used to make up the basic assembly-training film. The training procedures for the three groups thus differed only with respect to familiarization-training. In detail the following procedures were used:

1. No-Familiarization Procedure. Subjects trained in this way were shown only the basic assembly-training film before being tested. Total training time was about 22 minutes.

2. Prior-Familiarization Procedure. Prior to viewing the basic film, subjects trained by this procedure were familiarized with the actual parts of the distributor, which had been mounted on masonite boards. These parts were shown to the trainee one at a time in a predetermined order which was irrelevant to the assembly order and which was the same for all trainees. After each part had been shown once, all were shown again in a second irrelevant order. During the first showing, the experimenter pointed to the important features of each part. During the second showing each part was handed to the trainee and he was requested to point to the important features. Errors of omission were corrected by the experimenter. These two familiarization trials required about 24 minutes. Immediately following this familiarization the basic assembly-demonstration film was shown, making the total training time about 46 minutes.

3. Film-Familiarization Procedure. The only training given under this procedure was to show the film-familiarization version of the basic training film. As described above, the version was made up by supplementing the basic version of the film with "familiarization" close-up shots. The features pointed out in these close-ups were the same as those pointed out in the prior-familiarization procedure. Total training time for this procedure was about 30 minutes.

Administration and Subjects

A total of 33 experimental subjects was used, eleven under each condition of training. Each subject was trained and tested individually. Each was trained by one experimenter and tested immediately afterward by a different experimenter who did not know what condition of training had been used.

The subjects were prison inmates between the ages of 16 and 26. Only men with a Stanford-Binet IQ between 76 and 120 were used, with the additional restrictions that they had scored between the 25th and the 99th percentiles on the Minnesota Paper Form Board, and that they had only a minimum of experience with an automobile distributor.

The Minnesota Paper Form Board scores were obtained prior to the experiment and were used as a basis for matching subjects. The three subjects with the highest scores were run as one block, with one subject under each training condition; the next three highest were run as a second block, and so on. There were, therefore, eleven blocks of this kind.

Performance Test

All subjects were given the same test after instruction. In the test each subject was required completely to assemble a distributor, following a set order of sub-tasks. For this purpose the total assembly was broken down into four sub-tasks, and each of these was accomplished as a separate unit by the subject. For each sub-task he first was required to select all of the necessary parts from a superfluous array of parts arranged in a standardized pattern. Time and error scores were recorded for each sub-task selection. For example, the first sub-task included all of the parts that must be attached directly to the "main shaft." The main shaft was handed to the subject and he was told to select all of those parts. Both errors of omission and unnecessary parts selected were tallied, as well as the time from "start," until the subject indicated that he was satisfied with his selections.

After the selection for a sub-task, the experimenter corrected any errors and the subject was told to assemble the parts. Again time and error scores were tallied. An error was scored whenever a part was incorrectly assembled or an incorrect attempt to assemble a part was made. An error was also scored if the subject remained at an impasse for two minutes. More than one error per part was allowed if the subject corrected the error before going on to the next piece or before the end of the two-minute limit. Impasses and errors not corrected by the subject were corrected by the experimenter, but time spent in making these corrections was not scored. Within a sub-task the subject was free to assemble parts in any order which would not require the removal of a part in order to add subsequent parts.

Results

The major results are shown in Figures 2 and 4, derived from the tables which appear at the end of the chapter. Figure 2 shows graphically the mean number of selection errors for the three different training groups and for the two familiarization groups combined. It can be seen from the graph that both familiarization groups made fewer selection errors than the no-familiarization group and that the film-familiarization group made the fewest errors. When all 33 individual error scores are ranked without respect to training group, the mean rank for the no-familiarization group is 23.0, as compared with a mean rank for both familiarization groups combined of 14.0. The Mann-Whitney rank tables (Auble, 1953) show that a mean rank of 23.0 or more for the no-familiarization group would occur by chance less than one per cent of the time, and we may conclude that the familiarization groups performed reliably better than the no-familiarization groups with respect to selection errors. Moreover, when the two familiarization groups are compared by the Wilcoxon rank method for paired scores (1949), it is found that the Film-Familiarization group performed reliably better ($p = .03$) than the Prior-Familiarization group with respect to selection errors.[2]

[2]See Tables 1 and 2 at the end of the chapter.

Mean selection times are summarized in Figure 3. While the ranks of the relationships among the mean selection times shown in Figure 3 are the same as those for mean selection errors, none of the differences between groups is reliable.

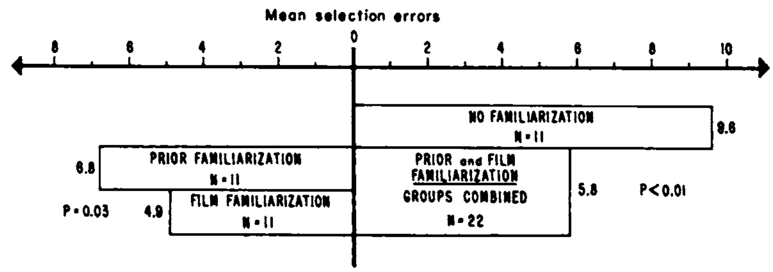


Figure 2. Mean selection errors for the No-Familiarization group, for both Familiarization groups combined, and for each Familiarization group separately

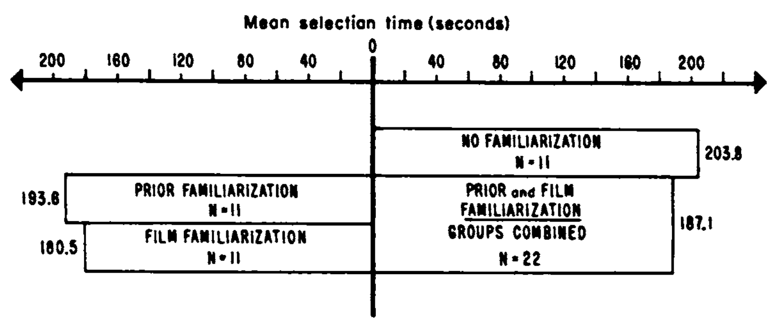


Figure 3. Mean selection times for the No-Familiarization group, for both Familiarization groups combined, and for each Familiarization group separately

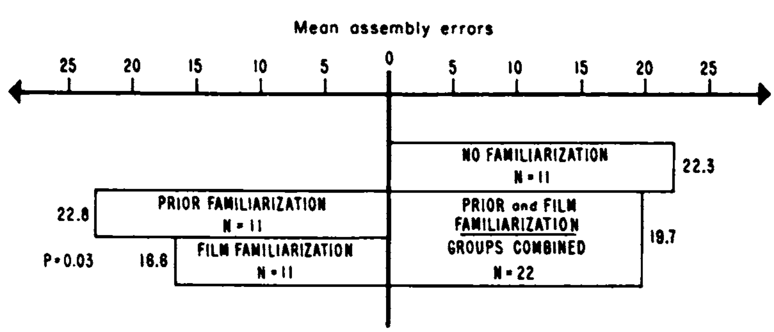


Figure 4. Mean assembly errors for the No-Familiarization group, for both Familiarization groups combined, and for each Familiarization group separately

Figure 4 shows graphically the mean number of assembly errors for the three different training groups. Although the direction of the differences with respect to assembly errors is the same as that for selection means, the combined Familiarization group was not found to differ reliably from the No-Familiarization group. However, a Wilcoxon rank test for paired scores indicates that the Film-Familiarization group made reliably fewer assembly errors than the No-Familiarization group. ($p = 0.03$)[3] Mean assembly times are shown in Figure 5. No reliable differences were found among the groups with respect to assembly time.

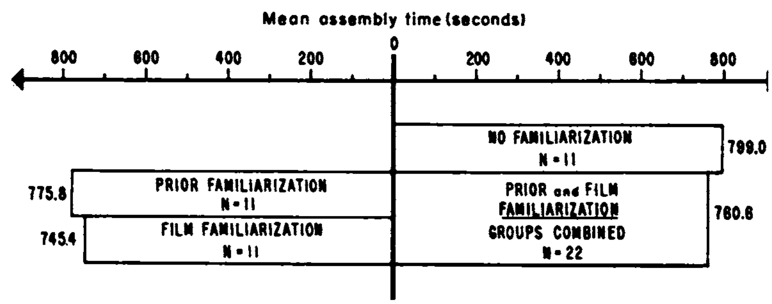


Figure 5. Mean assembly times for the No-Familiarization group, for both Familiarization groups combined, and for each Familiarization group separately

Discussion

The results for selection errors tend to support the expectation that familiarization by pointing to important features of the mechanical parts would facilitate part selection during the test assembly. These results also show that familiarization is superior when it is done for each part just before the assembly of the part is demonstrated—as compared with prior familiarization. The differences favoring Familiarization over No Familiarization and Film Familiarization over Prior Familiarization were statistically significant for selection errors; they were in the same direction for assembly errors and for the two subsidiary time measures, but none of these latter comparisons was found to be statistically significant.

The inference to be made from these findings about overall performance in the assembly of the distributor is not clear without careful consideration of the test procedures used in this experiment. For example, the effect on overall assembly time of the reduction in selection errors effected by familiarization would most likely be quite dramatic if the test assembly task were to assemble all of the parts from beginning to end without aid from the tester. In such a test situation, every wrong part selected would require extra trial-and-error time before being discarded in favor of another choice. Likewise, with such a practical test situation

[3] See Tables 3 and 4 at the end of the chapter.

assembly errors would increase. A structured test was used in this experiment because a free test situation is difficult to score, especially when separate scores for selection and assembly are desired.

When these factors are given consideration, it is apparent that the present results are strong evidence in favor of the practical utility of familiarization training for this mechanical-assembly task. The results also show that prior familiarization is inferior to familiarization given during assembly training, both as measured by error scores during testing and in terms of training time. It may be recalled that the Prior-Familiarization group required 45 minutes of training time as opposed to only 30 minutes for the Film-Familiarization group. The comparison between No Familiarization and Familiarization with respect to selection scores is also consonant with the expectation derived from the theoretical position presented. Thus, the superior test performance of the Familiarization groups may be attributed to more frequent attention to appropriate cues during assembly training. The failure to find the expected difference with respect to assembly errors may be due in part to the limited sensitivity of the test procedure used.

What are the more general implications of these findings? Certainly the evidence that is presented by this single study utilizing a quasi-practical training task as an experimental vehicle cannot be taken as more than suggestive evidence. However, the findings do suggest the utility of the underlying approach, which is to analyze tasks to determine the kind of learning problem that must be solved in training, and then to tailor appropriate training conditions on the basis of the analysis. The results also suggest that one kind of learning problem to look for in analyzing tasks is the problem of cue-utilization, and that one kind of training condition that is useful when a cue-utilization problem is present is some sort of "familiarization" procedure during training.

Summary

Using the assembly of an automobile-ignition distributor as a task, two training procedures with familiarization were compared with each other and with training without familiarization. One kind of familiarization involved training subjects to attend to important features of all mechanical parts prior to any assembly training; the other familiarization procedure involved pointing to the same important features of each part just before the assembly of each part was shown. Assembly training was accomplished by means of a 16-mm. black and white sound-movie film especially prepared for the experiment.

During subsequent testing of ability to assemble the entire mechanism, time as well as assembly errors and errors in the selection of parts were scored. The groups given familiarization-training outperformed the No-Familiarization group with respect to selection errors. In addition, the group given all its familiarization prior to assembly-training performed less well than the other Familiarization group with respect to selection errors, even though the latter procedure required less time. No reliable differences among the groups with respect to time scores or assembly scores during test assembly were found, but all differences were in the same direction as those for selection errors.

TABLE 1

Selection-Error Scores for Each Subject by Experimental Condition with Selection-Error Means and Mean Ranks for Each Condition

	No Familiarization	Prior Familiarization	Film Familiarization
	4	4	1
	12	4	3
	9	3	2
	9	12	6
	10	9	7
	8	10	9
	8	4	3
	15	6	0
	6	8	4
	9	6	7
	16	9	12
Totals	106	75	54
Mean Errors:	9.6	6.8	4.9
Mean Rank:	23.0	16.6	11.4

Note: Figure 2 is based on the data in this table.

TABLE 2

Selection-Time Scores for Each Subject by Experimental Conditions
With Assembly-Time Means and Mean Rank for Each Condition
(Time in seconds)

	No Familiarization	Prior Familiarization	Film Familiarization
	279	187	88
	125	106	72
	400	153	134
	171	318	186
	137	151	258
	399	225	128
	139	126	230
	127	161	273
	163	356	116
	152	187	345
	150	160	156
Totals	2242	2130	1986
Mean Rank:	16.64	15.73	18.64
Mean Time:	203.82	193.64	180.54

Note: Figure 3 is based on the data in this table.

TABLE 3

Assembly-Error Scores for Each Subject by Experimental Condition With Assembly-Error Means and Mean Rank for Each Condition

	No Familiarization	Prior Familiarization	Film Familiarization
	9	20	9
	15	2	4
	29	1	3
	17	74	32
	32	9	29
	12	14	16
	16	22	8
	23	14	15
	19	31	16
	47	37	29
	26	27	22
Totals	245	251	183
Mean Errors:	22.27	22.82	16.64
Mean Rank:	19.18	16.95	14.86

Note: Figure 4 is based on the data in this table.

TABLE 4

Assembly-Time Scores for Each Subject by Experimental Condition
With Assembly-Time Means and Mean Rank for Each Condition
(Time is in Seconds)

	No Familiarization	Prior Familiarization	Film Familiarization
	571	547	358
	484	267	262
	1164	250	304
	730	1420	1084
	775	602	1122
	800	722	1022
	696	706	631
	536	448	563
	599	1488	905
	1223	1020	888
	1211	1064	1061
Totals	8789	8534	8200
Mean Rank:	15.82	17.73	17.45
Mean Time:	799.00	775.82	745.45

Note: Figure 5 is based on the data in this table.

CHAPTER 11

FILM-MEDIATED LEARNING WITH VARYING REPRESENTATION OF THE TASK: VIEWING ANGLE, PORTRAYAL OF DEMONSTRATION, MOTION, AND STUDENT PARTICIPATION[1]

Sol M. Roshal

That the motion picture may contribute to perceptual-motor learning or the training of skills had been indicated, prior to this study, by the work of Hollis (1924), Lockhart (1944), McClusky and McClusky (1924), and VanderMeer (1945). More recently, studies by Jaspen (1949) and Zuckerman (1949) have supported this contention. However, many questions remain to be answered concerning how best to construct and use films for achieving the instructional objectives. It was the purpose of this study to develop one major hypothesis, along with four subsidiary hypotheses bearing on some related questions, and to report a set of experiments testing these hypotheses.

To promote clarity, certain terms used with various meanings should be rigorously defined for this discussion. These are: stimulus, response, perceptual-motor act, learning, and response strength. A stimulus is understood to mean any object or event complex which occasions an alteration in behavior. This usage does not imply simplicity or singularity of environmental detail. The singular form, stimulus, is defined to include the total situation which occasions the response. The plural form, stimuli, refers to repeated presentations of such a complex, or to more than one of such complexes. Neither is it implied that the stimulus necessarily elicits the response. It may simply serve to indicate the occasion for the response. Certain aspects of the stimulus may, however, be more important than others. Any stimulus-occasioned act is called a response. Whether or not there are or can be acts which are not stimulus-evoked is irrelevant here. The learning dealt with in this research requires a given response on the occurrence of a given stimulus.

[1]The study on which this chapter is based was conducted in 1948-49 as part of the Pennsylvania State College's Instructional Film Research Program, supported by a contract with the Office of Naval Research through the Naval Training Device Center (formerly, Special Devices Center). This program was under the general direction of C. R. Carpenter. The present report, prepared in relation to the author's subsequent role in the U. S. Air Force audio-visual research program (see preface) is a revision and condensation of the doctoral thesis by Roshal (1949b) and of a Technical Report prepared for the Navy Special Devices Center (1949a). See also Footnote 1, Chapter 20.

A perceptual-motor act refers to the overt moving or manipulation of some body part in response to a stimulus. Thus, the following of a moving object with the hand in a pursuit task or the placing of an object in a specified position would be examples of perceptual-motor acts. Learning implies the change in strength of a response to a stimulus in consequence of responding to the stimulus. The response strength may change for other reasons, and it might prove fruitful to examine these other variables in the context of instructional films; but the present concern is with learning as defined. Response strength is used here to mean the probability of occurrence of a response. It is measured (or inferred) as manifested in the number of responses.

Theoretical Considerations

Unless it is desired to set up a response to the film itself, the film never presents the actual stimulus to which a response is desired. Further, it is unreasonable to assume that the stimuli presented in the film will automatically evoke the exact response or responses desired. If this be so, it is necessary to make some assumptions about what happens in film-mediated learning in order to account for the occurrence of learning of the desired kind.

Assumptions

First, it is necessary to assume that there is some transfer from the stimulus used in training; that is, that the learning of a response to some stimulus will have some effect on the response to other stimuli. Second, it must be assumed that during the viewing of the film the learner does respond in some way to the stimuli presented. He may actually follow through the desired response overtly; he may make the response implicitly; or, he may make the response symbolically by verbalization or other self-signaling devices (cf. May, 1946). And third, if the actual desired response is not performed, we have to assume that there is some transfer to other responses from the response made in training.

The first assumption made may be stated more exactly, though still qualitatively, following Hull's "Principle of Stimulus Generalization" (1943, P. 183). During practice, a response is set up not only to stimulus used, but to a family of similar stimuli, the strength of the response to any particular stimulus being directly proportional to the similarity of the particular stimulus to the stimulus used in training. Two studies by Hovland (1937a, b) lend experimental support to this assumption.

In order to make this postulate quantitative, it would be necessary to state a function of similarity. Nevertheless, certain conditions for film production may be deduced without introducing quantitative rigor. It may be deduced that, for greatest efficiency of learning, the stimuli presented in the film should be as similar as possible to the stimuli which would occur in the actual performance (or test) of the desired learning. The film representation should approximate, as nearly as practicable, the details of the actual stimulus in terms of color, shading, perspective, sounds, content, etc. It should be expected, however, that for some stimuli the differences may be very small on a scale of similarity, and thus, for practical purposes inconsequential.

Implications

In a perceptual-motor act, whether or not there is a discrete sequence, the stimulus and the occasioned act are constantly changing. These changes may be of two kinds. First, there may be continuous changes, as in a pursuit task (see Chapters 12-15) or a radically changed configuration of the object acted upon, as in an assembly task (see Part I and Chapter 10). Second, there may be changes in the disposition of the musculature and body parts of the performer (e.g., the performer's hands). Such changes become the stimuli for further action. Thus, if the film-mediated learning situation is to represent the actual stimulus of the test situation, it must reproduce all these changes as exactly as possible. Hence the major hypothesis of this study: The learning of a perceptual-motor act, through films, will be more effective as the film presentation approaches a representation of the learner himself performing the act desired.[2]

The third assumption may be restated in an analogous fashion: During practice, a family of responses to the stimulus used is set up in addition to the practiced response, the response strength of any particular response being directly proportional to the similarity of the particular response to the response practiced.[3] It might simply be assumed (the second assumption above) that the learner will in some way respond to the stimuli. This postulate, however, goes further to imply that the film-learning situation should be so designed as to evoke or involve as close an approximation as possible of the actual responses of the test situation. For instance, for certain tasks it may be possible to have the learner perform the operation depicted by the film as the film is being shown.

Following the analysis of Gagné and Foster (1950), who separate the motor and perceptual aspects in their discussion, it may be that for a large number of tasks the movements involved are already a part of the learners repertoire and well practiced. The problem then becomes one of the adequacy of the perceptual aspects. In such a context perceptual adequacy might be considered the major concern of this paper (cf. Chapters 2 and 9 of this volume). If, in addition to the previously stated postulates, it were further assumed that the subject has learned to make certain responses on the occasion of perceiving the sensory consequences of such movements, it would be important that the film, within its limitations, approximate these sensory consequences by being as complete as possible.

[2]The reader may find it of interest to compare this hypothesis with the first of the principles presented by Sheffield and Maccoby in Chapter 9, and with some similar principles previously stated by May (1946, 1958).

[3]It would now be possible to derive a statement similar to Hull's third principle of generalization, stimulus-response generalization, but this is considered unnecessary.

Identification

The foregoing discussion has some implications for the problems of learner-protagonist identification in the film-instruction situation. Identification is here conceived in the sense of "acting along." The learner is said to identify with the film protagonist if he "acts along" with the actor in the film. Neglecting other dimensions of identification, it is suggested that one of the variables involved is the similarity between the protagonist and the learner. As the perceptual-motor activity proceeds, the learner perceives the sensory consequences of his action. The learner has previously experienced sensory consequences in terms of himself, and may have certain responses already learned. If the perception of sensory consequences can evoke the act, the most effective consequences would occur when the protagonist is most like the learner, thus approaching a representation of these sensory consequences.

Design of the Experiment

To test the validity of the major hypothesis of this research, four limited propositions, each of which could be translated into an appropriate film variable, were deduced. A simple perceptual-motor task—the tying of each of three standard knots—was selected, and eight versions of a film were produced demonstrating the tying of each of these knots. Each of the eight film versions involved a unique combination of one of the alternatives for each variable. These film versions were shown to groups of Navy recruit trainees, who were tested on the tying of each knot immediately after seeing the film demonstration of it.

The experimenter is always confronted with the possibility that a general hypothesis may be true but unsupported by his experiment because (a) the extension to the experimental hypothesis is illogical, (b) the particular situation selected is irrelevant, and/or (c) the experimental design employed is not sensitive enough to detect the differences in results that occur as a consequence of the experimental variables. A particularly crucial issue arises in the application of the concept of similarity, as it applies to the relationship between the film and the performance situation of which the film is a representation. Since similarity is not quantitatively defined, points on the continuum cannot be located definitely. This problem could be handled in one of three ways. First, some external criterion could be set up for scaling similarity, and points on a (more or less arbitrary) continuum could be identified. Second, similarity might be treated as an intervening variable. Third, one might try to choose points that are "obviously" far apart. The last method was followed in this experiment. The attendant difficulties require little elaboration.

Hypotheses

With these considerations in mind, the four limited experimental hypotheses and their associated film variables may be stated as follows:

1. Camera Angle. The first hypothesis is that film mediated perceptual-motor learning will be more effective as the film more accurately represents the visual stimuli as the learner would actually see them in performing the act.

An obvious consideration in the presentation of a perceptual-motor task is the point of view or perceptual orientation used. If the film situation is to approach a representation of the learner himself performing the perceptual-motor act, the screen presentation must show the act as nearly as possible as it will actually appear to the learner as performer. This means that the camera should be oriented with respect to the task as a learner would be oriented in performing the task, rather than as a learner watching an instructor in a classroom. The former orientation is known to professional motion-picture workers as the "subjective" camera angle, while the latter is called the "objective" or "observer" camera angle. In this study they have been termed the 0-degree camera angle and the 180-degree camera angle, respectively. There can be no doubt that two distant points on the similarity continuum are thus depicted. From the hypothesis it may be deduced that the films employing the 0-degree camera angle should be more effective than those employing the 180-degree angle.

2. Motion. The second hypothesis is that film-mediated perceptual-motor learning will be more effective as the film more accurately represents all the changes in the object acted upon.

This hypothesis is tested by a comparison of film versions which show continuous movements, or "live action," with static versions showing only successive stages in the knot-tying process. From the hypothesis it would be predicted that a version incorporating motion should be more effective than a static version.

3. Hands. The third hypothesis is that film-mediated perceptual-motor learning will be more effective as the film more accurately represents the disposition of the relevant body parts of the performer.

It has been pointed out in the previous chapter that the disposition of the body parts of the performer should be considered as part of the stimulus, and further, that changes in their disposition constitute stimuli for further action. A film, then, which shows the disposition of the body parts should be more effective than a film which does not. Since the relevant body parts for the task selected are the hands, the variable subsumed has been labeled "hands."

It had been intended to test this hypothesis by the comparison of an animated-film version which showed continuous motion without a human agent, versus a version showing live action including the hand movements. This plan was discarded because of the expense of animation, and, with some misgivings about their separation on the similarity continuum, a comparison of static versions was substituted. In two versions the hands are shown, and in two others, the sequence of steps in tying the knot in a line (rope) is shown in the film by showing the line without hands.

4. Participation. The fourth hypothesis is that film-mediated perceptual-motor learning will be more effective as learner response during the film presentation approaches the response required in the actual performance (or test).

It has been assumed that during the film presentation the learner does respond in some way. All of the films carried a sound-track explanation of the task, and response might be symbolic in terms of the verbal description. It is clear, however, that actually following through the desired response overtly is closer to actually performing the desired act in the test situation than any implicit response would be. If, then, the learner follows through the movements while watching the film, learning should be facilitated. Film versions which allow effective opportunities for such participation should be better than corresponding non-participation versions.

Experimental Films

Eight versions of a film demonstrating the tying of three knots were produced. The specific content of each version is presented in previous reports (Roshal, 1949a, b). In each version, each of the three knots was treated in the same fashion with respect to the film variables. Between each knot sequence was inserted a length of blank film, to warn the operator that the projector should be stopped for the test of learning on the preceding knot. Of the eight versions, four employed motion and four were static; six showed hands (all motion versions and two static versions) while two (static) versions did not. Six were photographed from the 0-degree camera angle and two from the 180-degree camera angle. Finally, four versions called for participation, while the other four, corresponding with respect to the previous three variables, did not. The distribution of variables in the eight film versions is presented in Table 1, and is also displayed, oriented in terms of the four experimental factors, in Table 2.

Task Characteristics

The following criteria were proposed to govern the selection of a task for this study:

1. It should be simple enough so that at least a portion of the tested population could perform it after one film demonstration, but difficult enough so that not everyone would be likely to succeed.

2. Correct performance of the task should not depend on an "insight" solution or the discovery of a "trick."

3. The task should be such that several specific instances of varying difficulty could be identified.

4. It should permit a relatively objective method of scoring, and should lend itself readily to testing. Preferably the task should permit testing rather large groups.

5. The task should incorporate elements that would permit the production of experimental film versions in which the variables previously described could be manipulated.

TABLE 1

Film Versions

Version Number	Version Code	Camera Angle (0^{o}--180^{o})		Motion (M--S)	"Hands" (H--NH)	Participation (P--NP)
I.	(180-M--H--NP)		180^{o}	Motion	Hands	No Participation
II.	(0--S-NH-NP)	0^{o}		Static	No Hands	No Participation
III.	(0--S-H--NP)	0^{o}		Static	Hands	No Participation
IV.	(0-M--H--NP)	0^{o}		Motion	Hands	No Participation
V.	(180-M--H--P)		180^{o}	Motion	Hands	Participation
VI.	(0--S-NH-P)	0^{o}		Static	No Hands	Participation
VII.	(0--S-H--P)	0^{o}		Static	Hands	Participation
VIII.	(0-M--H--P)	0^{o}		Motion	Hands	Participation

The tying of three knots, the bowline, the sheet bend, and the Spanish bowline, appeared to meet all the criteria stipulated. Since the study was to be conducted with Navy recruits, knot-tying had the additional advantage of face validity for the trainees. Three knots were used, to enhance the likelihood of a spread in performance in the results. For example, if all the subjects could tie a knot after seeing the film, independent of the particular versions viewed, no comparison could be made. The same would be true if none of the subjects succeeded in tying a knot.

It was also thought desirable to obtain an estimate, however indirect, of the interaction of the postulated variables with degree of difficulty. Direct assessment of the interaction would have required a complete confounding of the order of presentation of knots of known difficulty, and would have expanded the study to prohibitive proportions. A preliminary tryout of the task with college students indicated that one might anticipate a range of performance (from success to failure) for each knot. Furthermore, for this population the knots were of distinctly varying difficulty, the bowline being the easiest and the Spanish bowline the most difficult.

The task also lent itself well to group testing. Enough line could be provided to permit each subject to tie each knot, and the tied knots could be retained for subsequent scoring. However, while group testing presents definite advantages, the shortcomings of the procedure should not be neglected. When handling large groups, it is virtually impossible to make continuous observations of each subject's performance. Therefore, only the end-product can be scored.

If the purpose of the learning is to lead to the performance of any act to produce a particular product, no discussion of this point is necessary. If, on the other hand, the purpose is to teach a perceptual-motor act in all its aspects (i.e., a single "right" set of movements for tying each knot), then the scoring of the end-product only may lead to an attenuation of the data by inclusion of correct knots (end-products) tied by use of incorrect movements. This attenuation would probably work against the experimental hypotheses.

Experimental Procedure

About 4,200 recruits in companies of approximately 130 men each at the Great Lakes Naval Training Center, Great Lakes, Illinois, participated in the experiment. The use of several companies for exploring procedure, as well as wastage due to projector breakdowns and incomplete cases, resulted in an attrition of the experimental population to 3,314. Twenty-eight companies, one for each possible pairing of the eight films, were used. It was Station practice to form companies of men in the order of arrival. The men were all in their first week of training and had had no formal training in seamanship. Checks on informal training were made by interviewing the company commanders and samples of the recruits.

Materials

The eight film versions, each showing the tying of the bowline, the sheet bend, and the Spanish bowline, have already been described. The sections of Version I are a few seconds longer than the corresponding sections of the other versions. However, it is unlikely that this small discrepancy could account for any of the findings. In any case, it presumably should work against the camera-angle hypothesis.

Group testing required the preparation of many duplications of the test materials. Sufficient six-foot lengths of No. 7 sash cord were provided for testing at least two companies without reuse. The ends of the lines were cemented to prevent raveling. The lines were inserted in manila envelopes. Stamped on the face of each envelope, in large letters, was the name of the knot for which the included line was to be used.

Assignment of Film Versions

A completely symmetrical design permitting the comparison of each version with every other version was devised. In all, this design called for 28 pairs of film showings. Since each film version occurs seven times among the 28 combinations, each version was shown to one-half of each of seven companies. For example, Version I was shown to half of one company and Version II to the other half, Version VII was shown to half of a second company and Version II to the other half, etc.

It was the experimenters' intent to assign the film pairs to companies at random. The pairs were ordered I-II, I-III, ..., II-III, II-IV, ..., VII-VIII, and numbered serially. The order of the serial numbers was then changed in conformance with the table of random numbers given by Peatman (1947, p. 543-545). This order was followed except for two presentations, which had to be discarded because of projector breakdowns. These presentations were repeated after completion of the entire series. The complete distribution of companies by film version is given in previous reports (Roshal, 1949a, b).

Proctors

Proctors were provided by the Naval Training Station. The proctors assigned to the experimenters assisted in the handling of the materials, policing of the test situation, and scoring of the knots. The proctors were recruits who had completed several weeks of training. An attempt was made to select men who scored high on both the Navy General Classification Test and the Mechanical Aptitude Tests. A new group of proctors was supplied for each work week. The proctors were trained on the Saturday immediately preceding the week of service.

Test Procedure

Two test rooms were set up, each accommodating over sixty test positions. Before the test companies appeared, the proctors placed three envelopes, each containing a line, one for each knot, on alternate seats of the classrooms. For the participation versions, six envelopes were used. The men returned the lines to the same envelopes after attempting to tie the knots.

Each company was split in half at the classroom doors by assigning every other man in the marching order to each of the two rooms used. It was not possible to randomize the groups prior to their appearance. However, it was felt that the procedure used was adequate to obtain randomization.

After the subjects were seated, they were instructed in a more or less informal manner, as follows: "You are here this morning (afternoon) to take part in a special study of recruit training being conducted by the Office of Naval Research. You will be shown three short films. Each of these films will show you how to tie a knot. Immediately after the film, you will be asked to tie the knot. The films are very short, so you will have to stay right with them. In order to keep things moving in this room, you will not talk at any time, you will not handle the materials until I tell you to do so, and when I do, you will do so quickly."

The men entered identification information (Name, Serial Number, Company, Date) on the envelopes, and then the first film was shown. For participation versions, instructions were given to take the line from the appropriate envelope and tie the knot while it was being tied on the screen. (These instructions were repeated in the films.) Projection was handled by the regular seamanship instructors.

One film of each pair was presented to a random half (the A-half) of each company and the other film was presented to the second (B) half of the company. The two experimenters alternated in the supervision of the test situations for the several films. The projector was stopped after the showing of each knot. The men took lines from the appropriate envelopes and attempted to tie the knots within a two-minute time limit. These knots were later scored by the proctors, who had available reference photographs of the correct knots and several incorrect variants. Both practice and test knots were scored for the participation groups.

Analysis of Data

The knots were scored either right or wrong. (However, additional information on variants of the correct knots was collected for possible use in the event that the simple pass-fail criterion should prove too crude to detect differences among the versions even with the large population used.) Each individual, then, was scored right (assigned a value of "1") or wrong (assigned a value of "0") on each knot, and given a total score for the three knots. Counting "1" for each correct knot and "0" for each incorrect knot yielded a total score range of 0 to 3.

In order to compensate for the possibility that the differences in effectiveness among the film versions may have been too slight to appear in any striking fashion for individual learners, the 0-1 scores for each knot were summed over each experimental half-company. To increase sensitivity further, the effects were also summed over the three knots.

Predictions

In terms of the four hypotheses proposed for test, the relationships stated below should obtain among the versions. These relationships are also indicated in Table 2, in which the film versions are laid out in relation to the major experimental variables in such a manner that the predicted relationships can be rather directly perceived. In Table 2 the superiority postulated for the participation versions is indicated by the higher position of each such version (V, VI, VII and VIII) relative to the adjacent and otherwise identical non-participation version (I, II, III and IV). Also, the relative predicted status of pairs of versions defined by the other variables is likewise indicated for the adjacent pairs of versions—the 180-degree versions (I and V) being inferior, respectively, to the adjacent, otherwise similar zero-degree versions (IV and VIII), and the latter being predicted as superior, respectively, to III and VII, which are similar to IV and VIII except for the factor of motion. Finally, versions II and VI, the no-hands versions, are predicted to be inferior to versions III and VII, respectively. However, the relative standing of the non-adjacent pairs of versions in Table 2 is not necessarily of significance; e.g., the relative effectiveness of I or V versus II or VI would depend on the relative magnitude of effects due to the camera-angle factor as compared with those due to motion and hands, and this is not predicted by hypothesis.

More explicitly, the relationships predicted by hypothesis and represented in Table 2 are the following:

TABLE 2

Relationships Among Experimental Film Versions*

(1) Camera Angle:		180-degree	Zero-degree	
(2) Motion:		Moving	Moving	Static
(3) Hands	(4) Participation			
With Hands	P		VIII	
	NP		IV	
	P	V		VII
	NP	I		III
No Hands	P			VI
	NP			II

*Predicted relative superiority for adjacent versions, and adjacent pairs of versions, is indicated by the vertical position of the version-designating numeral (see text).

1. Camera Angle. Version IV (0-M--H-NP) should be more effective than Version I (180-M--H-NP); and Version VIII (0-M--H--P) should be more effective than Version V (180-M--H--P).

2. Motion. Version IV (0-M--H--NP) should be more effective than either Version II (0--S-NH-NP) or Version III (0--S-H--NP); and Version VIII (0-M--H--P) should be more effective than either VI (0--S-NH-P) or VII (0--S-H--P).

3. Hands. Version III (0--S-H--NP) should be more effective than II (0--S-NH-NP); and Version VII (0--S-H--P) should be more effective than VI (0--S-NH-P).

4. Participation. Versions V, VI, VII, and VIII should be more effective, respectively, than Versions I, II, III, and IV.

5. Overall Order. In terms of the major hypothesis, certain hierarchies of effectiveness may be predicted, in addition to the specific comparisons between pairs of versions indicated above. (These are contained more or less explicitly in the discussion of the four subsidiary hypotheses.) Version VIII, characterized by motion, hands, 0-degree angle, and participation, should be the most effective of all eight, while Version IV should be the most effective of the four non-participation

versions. Versions IV, III, and II should rank in that order of decreasing effectiveness, and similarly Versions VIII, VII, and VI.

A statistical analysis termed the "Random-Half Control Method" was developed for this study. The method is sensitive to the fact that, with the procedures for assignment to experimental treatments used, the sampling is by half-company, whereas randomization of individuals obtains only within a company.

Since each film version occurs seven times among the 28 combinations, seven randomly selected half-companies viewed each film, and seven replications for comparison of any pair of film versions can be constructed. There is first a direct comparison, using one company for each film pairing. By controlling on the other six film versions, six more indirect comparisons are made available. For example, Versions I and II were paired and presented to random halves of one company. The same was done for the pairs I and III, and II and III. By pivoting on film III, an indirect comparison of I and II can be made: (I - III) - (II - III') = I - II (where III and III' indicate two presentations of Version III). Of course, III is not necessarily equal to III', but it is assumed that, except for random sampling error, an estimate of which enters the standard-error formulation, any difference between III and III' reflects systematic differences in the populations which are thus controlled. It is further assumed that these systematic effects are additive and that, for present purposes, a difference of 10 per cent of correctly tied knots between 50 per cent and 60 per cent is equal to a 10 per cent difference between, say, 10 per cent and 20 per cent.

It would be possible to expand the number of comparisons by pivoting on two versions, but the advantages would probably not be worth the extension of the assumptions and increase in size of sampling errors. There are then seven comparisons of Versions I and II:

I - II
(I - III) - (II - III')
(I - IV) - (II - IV')
(I - V) - (II - V')
(I - VI) - (II - VI')
(I - VII) - (II - VII')
(I - VIII) - (II - VIII')

Similar sets of seven comparisons are available for all of the 28 film pairings.

Following through a set of seven comparisons for any given film pairing yields seven independent estimates of the probability of obtaining a difference as large as the one observed between the members of the pair. Since the probabilities are not additive, Lindquist (1940, p. 46-47) suggests converting them into Chi-square values, which are additive. The combined probability is the probability for the summed Chi-squares with, in this case, fourteen degrees of freedom.

Throughout this study, the implied null hypotheses assume a true difference of zero. Since the null hypothesis will here be rejected only when the difference is in a specified direction, the probabilities reported represent only one tail of the chance distribution.

Results

Results for "control" groups who were tested without seeing a film showed that, of 112 subjects, seven per cent could tie the bowline, approximately three per cent the sheet bend, and none the Spanish bowline, without having been given instruction by films or otherwise. No further consideration will be given to film instruction as compared with no instruction.

Table 3 gives the proportion of recruits who successfully tied each of the different knots following exclusive instruction by one of the experimental film versions. Also, in the last column the mean total score for all three knots for each version is given. The last line shows the proportion of knots tied without respect to film version.

TABLE 3

Number of Subjects, and Proportion of Subjects Successful in Tying Knots, for Each Film Version

Version Number	Version Code	Number of Subjects	Individual Knots: Bowline	Individual Knots: Sheet Bend	Individual Knots: Spanish Bowline	Mean Total Score
I.	(180-M--H--NP)	418	.256	.103	.110	.469
II.	(0--S-NH-NP)	428	.582	.103	.028	.713
III.	(0--S-H--NP)	417	.439	.082	.034	.554
IV.	(0-M--H--NP)	424	.542	.311	.255	1.108
V.	(180-M--H--P)	405	.254	.168	.042	.464
VI.	(0--S-NH-P)	397	.589	.123	.040	.753
VII.	(0--S-H--P)	409	.469	.073	.032	.575
VIII.	(0-M--H--P)	416	.587	.351	.255	1.192
All Versions		3314	.465	.165	.100	.730

Statistical analysis of the results for the experimental comparisons will first be presented for the total scores, summed for all three knots. Results for individual knots will then be considered.

Total-Score Analysis

The twenty-eight possible comparisons among the total scores for the eight versions are presented in Table 4 in terms of the Chi-square values, derived as explained above. These comparisons provided a test of the experimental hypothesis with respect to knot-tying performance independent of task difficulty.

TABLE 4

Total Score: Chi-Square Values for Random-Half Control Analysis

Film Version	Film Version II	III	IV	V	VI	VII	VIII
I	-41*	+	-120*	+	-39*	-	-155*
II		43*	-73*	50*	+	34*	-94*
III			-120*	+	-36*	-	-150*
IV				124*	68*	116*	-26
V					-34*	-	-157*
VI						34*	-89*
VII							-145*

Sign of entry: Negative indicates that the higher-numbered film version (top) is superior to the lower-numbered version (side); positive indicates lower-numbered film version is superior.

*Indicates 1 chance (or less) in 100 of getting differences as large or larger in the hypothesized direction if the true difference were 0 and only chance factors were operating.

Absence of the asterisk indicates 5 chances (or less, but more than 1) in 100.

"+"without a numerical value indicates positive, and "-" indicates negative Chi-square values for which the chances are greater than 5 in 100..

Any interactions which might be suggested by other analyses are obscured in this total-score analysis. Explication of the major results may also be aided by referring to Figure 1, which shows the comparisons among the total scores for the eight versions.

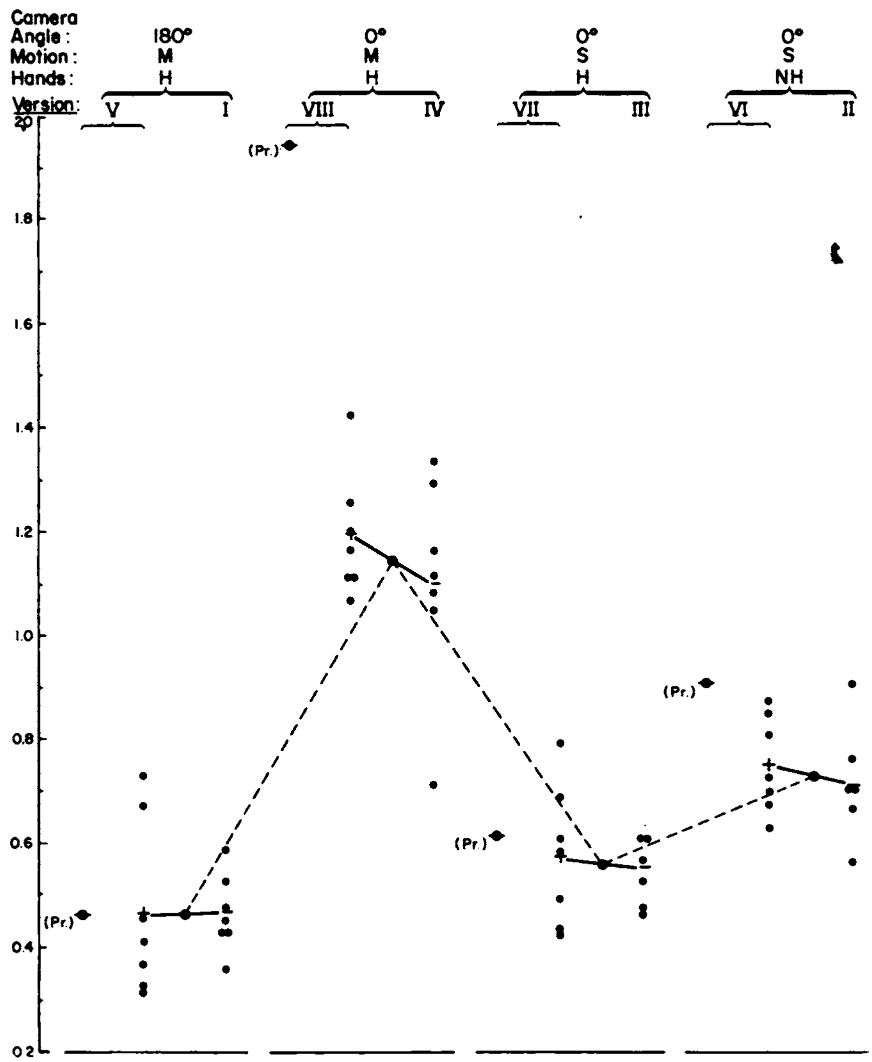


Figure 1. Total-score means (combined performance on three knots), with averaged means (large dots) for pairs of versions with participation (+) and non-participation (-).

Small dots above and below version means indicate scores for individual half-companies.

Unconnected large dots with line through them (labeled "Pr."), indicate means for practice for that version.

These have been plotted in a sequence to facilitate comparison with the predicted relationships summarized previously in Table 2. The means for each adjacent pair of corresponding participation and non-participation versions, labeled "+" and "-", respectively, are also shown in relation to the mean number of practice knots tied during the participation versions. The distribution of total test scores for each of the now half-companies is shown by the dots above and below the means for each version. Inspection of the degree of overlap of these distributions, as well as the Chi-squares given in Table 4, is indicative of the significance of the differences between versions or pairs of versions. Finally, since the participation versions generally did not differ significantly from the corresponding non-participation versions (see below), the means of the means for pairs of versions differing with respect to the other three experimental variables are plotted to show graphically the overall differences associated with each variable. Referring both to data plotted in Figure 1 and to the Chi-square values in Table 4, the following findings are revealed concerning the effects of the four experimental variables:

1. Camera Angle. The superiority of the 0-degree camera angle is unequivocal. Both versions IV (0-M-H-NP) and VII (0-M-H-P) are obviously superior to their counterparts, versions I (180-M-H-NP) and V (180-M-H-P).

2. Motion. The basic comparisons for this variable are films IV (0-M-H-NP) versus III (0-S-H-NP), and VIII (0-M-H-P) versus VII (0-S-H-P). All these films include hands, but III and VII are static. The major difference is the movement. It can be stated with great confidence that the films portraying continuous changes through movement are the more effective films. It would seem that the experimental hypothesis is supported.

3. Hands. Comparison of static hand versions III (0-S-H-NP) and VII (0-S-NH-P) with the static no-hand versions, II (0-S-NH-NP) and VI (0-S-NH-P), favors the films which do not portray hands. While the inclusion of hands was not expected to be a large step along the realism continuum, especially in view of the static portrayal, neither was it anticipated that this addition would weigh in the opposite direction. Two alternate hypotheses, which should be considered in further work to resolve this discrepancy, are suggested:

a. The inclusion of hands must, perforce, obscure some of the line, thus reducing the accuracy of the presentation of the object acted upon or the clearness of the significant perceptual fields. If the increased accuracy of representation of the performer is smaller than the decrease in the representation of the object, the resultant is not in the desired direction. It would be necessary to find some technique for weighing these factors.

b. Inspection of the films themselves leads to a second consideration. Dissolves were used between stages of tying the knots. It may be that dissolves in the versions without hands may have conveyed transitional relationships between stages, while, when hands were shown during the dissolves, the composition of the pictures became crowded and confused. Relationships could not be clearly perceived.

It might also be pointed out, as will be discussed later, that the major contribution to these findings is made by the bowline. The first point discussed immediately above may be pertinent particularly for this knot.

4. Participation. When the first three versions are compared with the corresponding participation presentations, the differences found are best explained in terms of chance. Version VIII, however, can be said to be better than version IV with a moderate degree of confidence.

The equivocal nature of this outcome may be attributed to the fact that participation as defined was not actually achieved. While all subjects attempted to "participate," a proportion never succeeded in tying knots during the film showing. It may be seen in Table 5 that, except on Version VIII, only a few more knots were

TABLE 5

Number of Practice and Test Knots Tied for Participation-Film Versions

			Individual Knots							
			Bowline		Sheet Bend		Spanish Bowline		All Knots	
Version Number	Version Code	N	Practice	Test	Practice	Test	Practice	Test	Practice	Test
V.	(180-M--H--P)	405	93	103	75	68	20	17	188	188
VI.	(0--S-NH-P)	397	261	234	76	49	24	16	361	299
VII.	(0--S-H--P)	409	214	192	31	30	8	13	253	235
VIII.	(0-M--H--P)	416	300	244	284	146	225	106	809	496
TOTALS		1627	868	773	466	293	277	152	1611	1218

tied during participation in the film viewing than during the test period. Also, the order of the number of knots tied during practice follows the order of effectiveness of the respective non-participation versions rather closely and the difficulty of the task inversely. It would seem, then, that in order to study this variable more closely, films must be so produced as to make participation possible; that this is more difficult for complicated tasks, and that as this objective is approached the contribution of participation may be realized.

5. Overall Order of the Versions. Although some of the positions in the hierarchy are not reliably determined, the several versions may be ranked in the following order of effectiveness with respect to total score: VIII, IV, VI, II, VII, III, I, and V. This is in agreement with the hierarchies predicted, except for the inversion of the hands and no-hands static versions.

Analysis by Individual Knots

With the exception of the placement of versions II (0-S-NH-NP) and VI (0-S-NH-P) higher in order of effectiveness for the bowline than for the other two knots, the findings for the individual knots are essentially consistent with one another and with the findings for the total scores. The differences among the versions in proportions succeeding in tying each knot are given in Table 6, and the test-means for each version for the individual knots are graphed in Figure 2, using the same arrangement as for Figure 1.

It should be recalled that the comparisons based on individual knots are not as sensitive as those based on the total score. The lack of refutation of the null hypothesis in these comparisons should therefore not be considered as refutations of the experimental hypotheses. Further, it is futile to discuss sign reversals in cases where the null hypothesis is not rejected, for the possibility of differences in either direction is predicted by this hypothesis.

The static no-hands version is particularly effective in teaching the tying of the bowline. This was not true for the other two knots. It may well be that in this version the photography for the three knots was not equally clear.

Two other possible explanations seem more reasonable. It may be that for a simple task it is only necessary to present a good picture of the thing acted upon, and the learner can readily supply the necessary movements from his repertoire.[4]

If the subjects were instructed to draw a "+", any number of possible presentations would be equally effective. It is doubtful that showing the necessary pencil movements or the development of the "+" would contribute anything. A slightly more complicated cross, say "ƚ", might require better presentation, but again, the main consideration would probably be the clarity of the cross itself. The no-hands versions, II and VI, showed the knots without any superfluous or obscuring detail.

An alternate or additional hypothesis might be that, given a simple task and ample time, the subject could search his response repertoire to find the proper movements if he had a clear percept. The time-limit used was two minutes. This is more than enough time to tie the bowline more than once. If the time-limit were reduced, less "searching time" would be available and the movements learned from the film should be more evident.

Comparison of bowline Versions II and IV were repeated for two companies, with a time-limit of 75 seconds. The change in favor of film IV was inconclusive,

---

[4] The availability of a repertoire of movements is irrelevant unless it is assumed that the stimuli presented can evoke the desired movements specifically, or it is assumed that only the end-product is of interest. If these assumptions are untenable, the question of attenuation due to scoring only the resultant must again be raised.

TABLE 6

Chi-Square Values for Random-Half Control Analysis, Individual Knots: (A) Bowline, (B) Sheet Bend, (C) Spanish Bowline

Film Version	Knot	Film Version II	III	IV	V	VI	VII	VIII
I	A	-93*	-37*	-60*	-	-83*	-41*	-101*
	B	+	+	-78*	-41*	+	+	-102*
	C	33*	39*	-62*	31*	36*	40*	-54*
II	A		46*	29	108*	+	45*	-
	B		+	-82*	-41*	+	-	-98*
	C		+	-111*	-	-	+	-113*
III	A			-31*	40*	-39*	-36*	-48*
	B			-98*	-48*	-	-	-124*
	C			-120*	-25	-	-	-109*
IV	A				66*	-	73*	-27
	B				44*	79*	85*	-
	C				100*	120*	121*	+
V	A					-90*	-44*	-109*
	B					38*	40*	-57*
	C					+	94*	-96*
VI	A						39*	-
	B						+	-94*
	C						+	-110*
VII	A							-49*
	B							-102*
	C							-113*

Sign of entry: Negative indicates that the higher-numbered film version (top) is superior to the lower-numbered version (side); positive indicates lower-numbered film version is superior.

*Indicates one chance (or less) in 100 of getting differences as large or larger in the hypothesized direction if the true difference were zero and only chance factors were operating.

Absence of the asterisk indicates five chances (or less, but more than one) in 100.

"+"without a numerical value indicates positive, and "-" indicates negative chi-square values for which the chances are greater than five in 100.

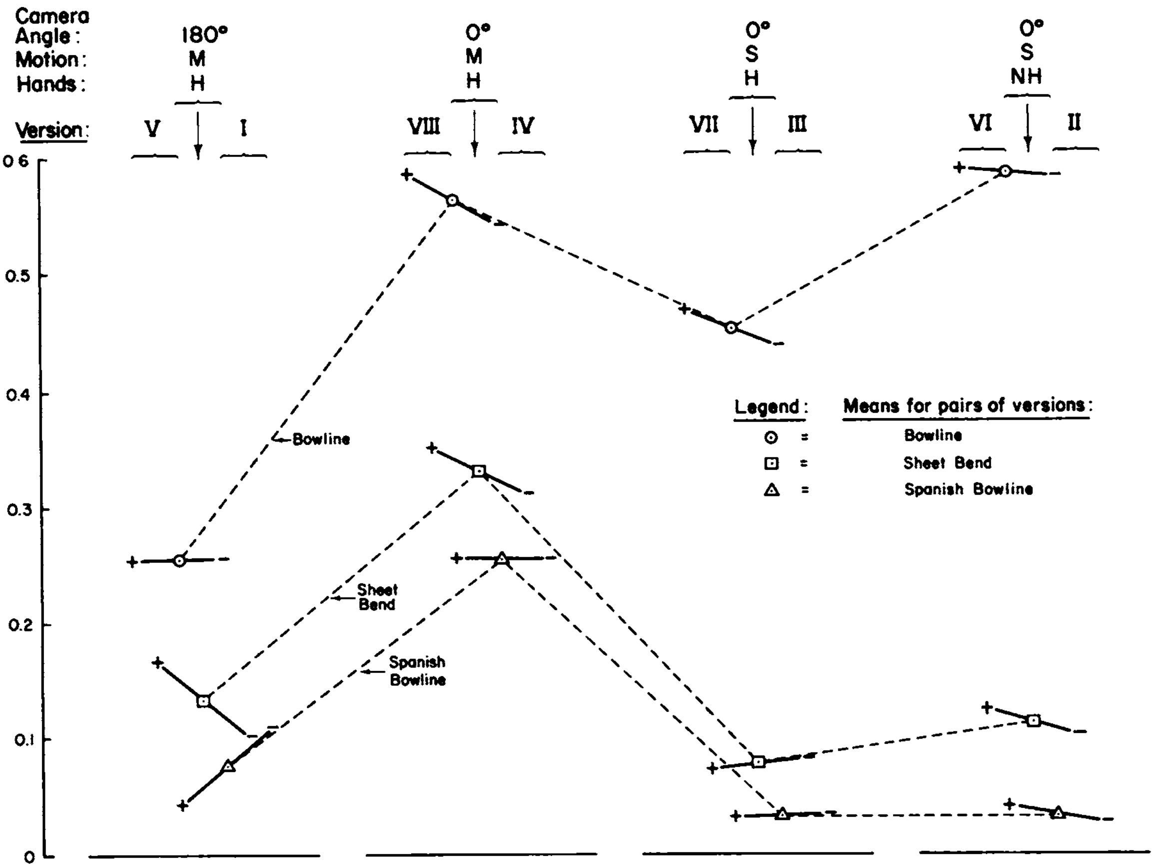


Figure 2. Individual-knot means for each version and for pairs of versions with participation (+) and non-participation (-).

but both of the above hypotheses are supported by the apparent greater importance of the motion-variable with increase in difficulty of the task. The differences between the several static versions and Versions IV and VIII tend to increase with task-difficulty.

Summary of Results and Conclusions

1. Films will be more effective in the teaching of perceptual-motor acts as they approach an accurate representation, continuously, of all the changes involved in the act. This includes changes in both the object acted upon and the relevant body-parts of the actor. That is, the film should show true movement and not simply static stages of the act. This finding probably does not hold for very simple tasks.

2. Films will be more effective when the perceptual-motor act to be learned is portrayed from a viewing angle approaching that of the learner actually performing the act. The subjective or 0-degree camera angle is more effective than the observer or 180-degree camera angle.

3. There was only limited support for the expectation that these film presentations would be more effective when an attempt was made to assure that the learner practiced responses approaching those involved in the actual performance of the act. Techniques should be developed to foster more effective learner-participation during film-viewing.

4. For simple tasks, the most important consideration may be the accurate representation of the product of the perceptual-motor act.

5. In general, learning of a perceptual-motor act through films will be more effective as the film presentation approaches a representation of the learner himself performing the act desired.

CHAPTER 12

SOME DELETERIOUS EFFECTS ON A PERCEPTUAL-MOTOR SKILL PRODUCED BY AN INSTRUCTIONAL FILM: MASSING EFFECTS, INTERFERENCE, AND ANXIETY[1]

William J. McGuire

Even passive viewing of a demonstrational film has been shown to increase the observer's manual skill at such motor tasks as typing (Sengbush, 1933), operating industrial machinery (VanderMeer, 1945), knot-tying (Roshal, 1949a,b), and mechanical assembly (Jaspen, 1950a. See also Chapters 1-9, *supra*). Such studies have usually been concerned with demonstrating the mean difference in performance scores between film and no-film conditions, or among various film conditions.

The present experiment attempted a more molecular evaluation of the effects of instructional films on perceptual-motor tasks by introducing two additional features. In the first place, it employed the rotary pursuit task on which the shape of the performance curves has been described (Ammons, 1947a), thus permitting comparison of theoretically interesting parameters other than mean performance-levels. Secondly, the film instructions were content-analyzed, and the Subject (*S*) was scored on how well he carried out the specific instructions as well as on how high a pursuit score he obtained.

The conditions of the present study and the response variables measured were selected to detect several possible detrimental effects that the film-viewing could have on subsequent performance. In the first place, it has been demonstrated (e.g., VanderMeer, 1945) that, aside from any mean effect, a film may decrease individual differences in motor skill. Particularly where the mean performance is not considerably raised, such a leveling effect suggests that potentially superior *S*'s were

[1]This study and those reported in Chapters 13 and 14 were conducted at Yale University under Air Force contract AF 18(600)-35, monitored by A. A. Lumsdaine for the audio-visual research division of the Human Factors Operations Research Laboratories (HFORL). Mark A. May was the project director, and F. D. Sheffield served as principal consultant. Several people contributed generously to various phases of this research, including production of the training film, apparatus construction, design of the experiment, and writing of the report. The writer, who is now at Columbia University, particularly acknowledges the aid extended by G. A. Kimble, J. J. Wulff, and F. D. Sheffield. The present report is a revision of a mimeographed Laboratory Note (McGuire, 1955a) which was based on a draft originally prepared in October, 1952.

reduced to a more mediocre level of performance (perhaps because the explicit and detailed visual demonstration in the film produced stereotyped responding and discouraged experimentation and adoption of idiosyncratic but effective techniques). Secondly, it was hypothesized that the rapid alteration of lighting conditions from film-viewing to performance could have a detrimental effect on skill when the performance followed the film-showing immediately. Third, on the basis of the "implicit-practice" theory (May, 1946; see also Chapters 2 and 11, supra.) it was hypothesized that seeing a demonstrational film immediately before performance of the task would tend to develop reactive inhibition of mediating implicit responses and thus, in effect, force the S to perform under relatively "massed-practice" conditions, with the usual work decrement resulting. Fourth, it was hypothesized that viewing the film would tend to increase the perceived importance of the task and thus raise the S's test anxiety, which is particularly likely to be detrimental to performance with this type of task. It was expected that even though the net effect of the film on motor performance might be favorable, analyses of other parameters of the practice curve and of other response variables might indicate that some of these detrimental effects were also operative and were subtracting from the gross benefits of seeing the film.

Method

Apparatus

A modified Koerth pursuit rotor was used. The turntable rotated in a clockwise direction at the rate of one revolution per second. The diameter of the target was 11/16 inch, and the diameter of the circle it described was 6.2 inches. The S's score was recorded on two Springfield timers that alternately recorded the pursuit score (i.e., time that the stylus was on the target) during successive 30-second intervals, permitting a fairly continuous record of the S's performance, since a reading on one clock could be taken and that clock reset while the other clock was recording.

Instructional Film

The S's in the film conditions were shown a film that contained a series of twelve instructions on how S should perform the pusuit task, similar to those listed near the beginning of Chapter 13. These instructions were illustrated in the film by showing an S performing the task. Each of the specific instructional items was demonstrated visually and described also on the accompanying sound-narration. The film was made so that there were three parts of lengths five, four, and four minutes, each of which contained the complete list of instructions and could be shown separately from the other two parts.

Procedure

Three practice sessions were given to all S's, each five minutes in length and separated by five-minute rest intervals. S's were run in one of two conditions: a film group which was shown the first part of the film before the first practice session, and the second and third parts before the second and third practice sessions; and a no-film group which "rested" (or engaged in task-irrelevant activity) during the inter-session intervals.

Subjects in the no-film group were instructed at the beginning of the first session as follows: "This rotor will turn in a clockwise direction at the rate of one revolution per second. You are to take this stylus (E points to the stylus) in your preferred hand and try to keep the point of the stylus (indicate) in contact with this silver target-disk (indicate) as much of the time as possible while the rotor is turning. Start trying to keep contact between the point of the stylus and the target-disk as soon as the turntable starts to rotate, and continue trying until the turntable stops. You will have three five-minute practice sessions separated by five-minute rest periods." The S's in the film group were given the additional instructions: "... Before you start now, and in each of the other five-minute rest periods between practice sessions, I shall show you a training film which demonstrates the techniques that have been found to result in high scores on pursuit tasks like these."

The scoring for each practice session was divided into 10 thirty-second "trials," and pursuit scores were recorded for each of these trials on alternate timers. From the S's point of view, however, each session consisted of five uninterrupted minutes of practice since the scoring was switched by the experimenter from one clock to the other while S's practice on the turning rotor continued uninterrupted. The S's were also scored on the extent to which they complied with the postural-adjustment directions advocated by the portrayal and narration in the film. During each of the three practice sessions each S was rated by the judge unfamiliar with the hypotheses on a three-point scale concerning the extent to which the S was using the postural adjustments advocated by the film. To the extent that the ratings were reliable this yielded data which could test separately the extent to which the instructions were followed as well as the extent to which following the instructions led to a higher pursuit score.

Subjects

The S's were 21 student nurses from a local hospital, who volunteered to serve as subjects in their off-duty hours and were paid one dollar each for their time. They fell into the 18-25-year age group and all had college training which included at least one course in psychology for nurses. The data on one subject had to be rejected because one of the Springfield timers ceased recording during a practice session. These S's were run over a six-day period at hours varying from 9 A.M. to 9 P.M., depending on their choice of time. S's were assigned to one or the other conditions according to a predesignated order.

Results

Mean Level of Performance

No obvious benefits from seeing the film were reflected in the mean pursuit-scores. In fact, the no-film group obtained a slightly higher mean total-contact time than did the film group (24.10 seconds vs 21.57). This difference fell short of conventional significance levels, though it was fairly pronounced throughout the second practice section (see Figure 1). On the other hand the film had a pronounced influence on performance which was reflected in postural scores. The mean ratings on utilization of the 12 recommended postural adjustments were 5.25 points higher in the film than in the no-film group ($p < .01$).

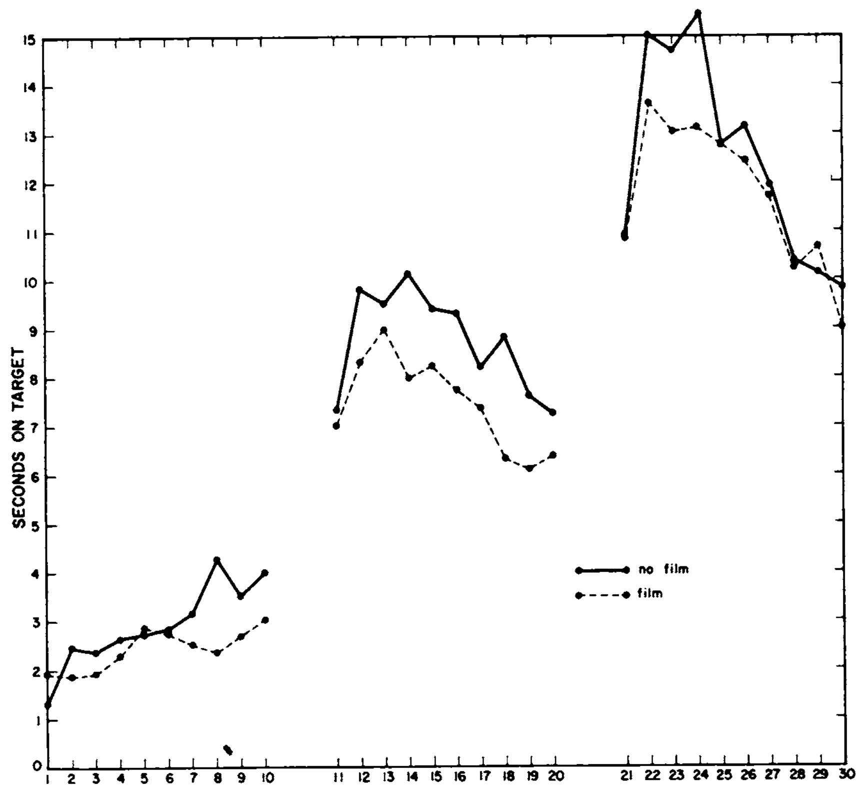


Figure 1. Pursuit scores: 30-second trials (five-minute session with five-minute rest)

The fact that the film group was clearly superior in performing in the way advocated in the film, but did not obtain superior pursuit-scores, suggests that the instructions advocated in the film were not in fact conducive to higher pursuit-scores. Further analysis bears out this suggestion. The overall coefficient of correlation between pursuit- and posture-scores is only + 0.12, and even an inter-class correlation within the two conditions (film and no-film) falls far short of conventional significance levels.

While the total postural scores does not show any appreciable relation to the pursuit-scores, we might ask the further question whether postural score on any of the individual items is significantly related to the pursuit-scores. Only three of the 12 advocated postural techniques showed any considerable relationship to pursuit-scores, namely, items (2), gripping the handle of the stylus with all four fingers and thumb, (9) relaxing, and (10) following the target continuously.

However, even though overall compliance with the filmed instruction did not notably affect the mean pursuit-scores, seeing the film can be shown to have affected two aspects of the pursuit-scores, namely, their variability and the shapes of the practice curves.

Variability Within Film and No-Film Conditions

While seeing the film had little effect on the mean level of the pursuit-scores, it did affect the variability of these scores. There is evidence that seeing the film reduced variability of the pursuit-scores in a number of respects, particularly in the first practice session. This reduction indicates that the potentially poorer performers were helped and the potentially superior performers hindered by seeing the film, since the difference in means was trivial.

Variability (in terms of standard deviation of scores) among the S's in the total pursuit-score for the three sessions was reduced from 10.87 seconds in the no-film group to 7.71 seconds in the film group. However, F-ratio of this difference is only 1.99, not significant even at the .10 level. When the total scores are broken down by session, the variability among subjects on the total pursuit-scores on each session was consistently lower for the film group, the between-group differences in standard deviations being 1.47, 1.41, and 0.19 seconds, for the first, second, and third sessions respectively. The first of these is significant at the .05 level. Besides this lesser inter-subject variability, the film group also shows less variability among the 10 mean trial-scores within each of the three practice sessions. The differences for the three sessions were .40, .09, and .69, though here again only the difference for the first session approaches significance (.10 level). Since all six of the standard deviations calculated are smaller for the film than the no-film conditions, we might conclude that seeing the film reduces interpersonal variance in pursuit-scores, though, as for the above differences, this reduction occurs primarily in the early stages of practice. This decrease in individual differences in motor skills after seeing training films has also been reported in connection with performance on industrial machinery (VanderMeer, 1945).

A possible mechanism of this leveling effect of the film, reducing the pursuit-scores of the potentially superior S's and raising that of the potentially poor ones, is that the film-condition S's showed a much more stereotyped mode of performance, as measured by the postural-adjustment scores. The standard deviations in total postural-adjustment scores for film and no-film groups were 2.53 and 4.70 respectively ($t = 2.02$, significant at the .06 level). These data suggest that by discouraging idiosyncratic modes of performance, the film-viewing eliminated individualistic approaches that tended to be particularly beneficial or detrimental to pursuit-performance. The above-reported finding that there was little relation between these postural and pursuit-scores does not necessarily nullify this interpretation of the leveling effect of the film, since presumably postural adjustments other than those selected for measurement were also affected by the film.

Practice Curves of Film and No-Film Groups

Two hypothetical detrimental effects of film-viewing would tend to show up in the shapes of the learning curve. If, as hypothesized, the rapid change in illumination level between the film-viewing and pursuit-performance conditions adversely affected the tracking score, this effect should be particularly pronounced in the early part of each practice session, when light adaptation was occurring most rapidly. The relevant data offer no support for this hypothesis; on the contrary, it is during the very first 30-second practice interval of each of the three sessions that the film group's performance tends to be best relative to that of the no-film group (see Figure 1).

The other hypothesized detrimental effect of the film that would be reflected in the shape of the practice curve has to do with massed versus distributed practice. It was hypothesized that devoting the five-minute intervals between practice sessions to viewing a film depicting the same task would constitute a more massed-practice situation than would obtain when the five-minute intervals were filled with task-irrelevant activities, as was the case in the no-film condition. Such an hypothesis would follow from the "vicarious-practice" theory (May, 1946) of the training efficacy of an instructional film on a motor-skill task. Assuming that the perceptual or ideational responses that play some mediating role in tracking behavior are occurring during film-viewing as well as during overt performance, and that these responses follow the same laws as gross motor responses in the development of reactive inhibition, we would expect the practice curves for the film group to resemble those of massed-practice conditions more than would those of the no-film group.

Assuming the no-film group had the longer effective inter-sessional rest, we can derive several predictions from Ammon's (1947a) theoretical analysis of motor-skill curves. Ammon's assumption 5, confirmed by his later (1947b) data from the same task (the pursuit rotor) as used in the present study, implies that there would be more warm-up (gain from the first to second 30-second trial of any session) in the no-film than in the film group. As can be seen in Figure 1, the no-film group does show a considerably greater warm-up effect in each of the three practice sessions. The overall difference in warm-up between the two conditions for all three sessions is significant at the .05 level (Table 1).

A further derivation from the distribution hypothesis and Ammon's deduction 5a (as well as Ammon's [1947b] data) is that the peak performance level during any session is reached earlier under the film conditions than in no-film conditions. This prediction also is borne out in all three of the practice sessions (see Figure 1).

Anxiety Levels in Film and No-Film Conditions

A third hypothesized detrimental effect of the film is that it might increase the anxiety level of the S's beyond the optimal point. A higher anxiety level might, of course, be beneficial to performance under some conditions (Lazarus, Deese, and Osler, 1952) but it was felt that the effect was likely to be adverse under the present conditions, considering that the paid volunteer nurses were probably already adequately motivated to do well, and that the task was one particularly likely to be adversely affected by the unsteadiness and interfering responses associated with increased anxiety.

TABLE 1

Analysis of Variance for the Warm-up Effect (Improvement on Second Trial over First, For Each of the Three Sessions, and For Film Versus No-Film Conditions)

Source of Variation	Degrees of Freedom	Sum of Squares	Variance	F	p
Condition (film, no-film)	1	24.58	24.58	5.74	<.05
Sessions	2	81.91	40.96	9.57	<.01
Interaction	2	0.22	0.11	0.03	>.05
Error	54	231.23	4.28	-	-

Some evidence that would tend to substantiate the prediction of greater anxiety under the film condition comes from an analysis of one of the postural-rating items that involved a rating of how relaxed the S was in each of the three sessions. This rating was made on a three-point scale by a paid judge unfamiliar with the hypotheses. The no-film group was rated significantly more relaxed ($p < .05$) than the film group on the basis of the composite three-sessions score. The further link in the argument—that anxiety is detrimental to pursuit-score—receives some support from the fact (mentioned in a previous section of the paper) that this "relaxation" item is one of the three out of 12 postural-rating items that showed any considerable relationship to the pursuit-scores.

Discussion

The present study succeeded almost too well in demonstrating detrimental effects of film-viewing on perceptual-motor skill. It had been anticipated that the net effect of the film on pursuit-scores would be beneficial, with the detrimental effects only partially offsetting the gross benefits and detected only by an analysis of some subsidiary response variables. As it turned out, not only were these offsetting effects manifested in the subsidiary responses, but no significant net beneficial effect of the film was found.

The trivial and possibly negative net effect of the film in this instance is not in itself of much theoretical interest. Two artifacts of the experimental procedures could easily account for the lack of beneficial film-effects in this case. In the first place, to increase statistical sensitivity, the experiment was deliberately designed to maximize these detrimental effects: in addition to the fact that the rotary-pursuit task maximizes work decrement and interference due to anxiety-provoked responses, distributed practice employed in the no-film condition was pitted against showing the film between practice sessions, without appreciable rest sessions, for the film condition. This would maximize any massing effect for the film condition. Also, the fact that performance followed immediately upon viewing would maximize any visual difficulties as well as the massing effect. Secondly, it happened (unintentionally) that most of the instructions explicitly portrayed in the film were not particularly conducive to a good pursuit-score. Hence, in terms of practical application, the magnitude of the detrimental effects of the film were probably atypical.

Nevertheless the detrimental effects of the film that were demonstrated here would tend to be operative even in situations in which the conditions were less loaded against the film and in which the net effect of the film was positive, tending to lower the magnitude of this net positive effect. Hence the problem of eliminating such detrimental effects is a general one.

It might be possible, for example, to correct the tendency of the film to produce uncreative imitation of the depicted technique (which depresses the achievement of potentially superior S's to a more mediocre level) by allowing an initial practice session before the film-showing. In this way, particularly rewarding idiosyncratic techniques can be tried and learned before the film discourages such venturesomeness. Since, as shown in the present study, the film does produce a massing effect when practice follows immediately after viewing, it seems advisable to introduce rest periods between viewing and performance. In this way not only might the reactive inhibition be reduced, but also, more important, one might reduce the more permanent conditioned inhibition consequent upon this reactive inhibition (Kimble 1949). Finally, the evidence that use of an instructional film can produce an undesirable increment in the S's anxiety suggests that some care be taken in programming the film to prevent this effect.

Summary

It was hypothesized that any net beneficial effect on motor skill produced by an instructional film would be partially offset by underlying detrimental effects which would be discernible by means of more molecular response analyses. An instructional film which presented (visually and in the accompanying narration) twelve directions for performing on the pursuit-rotor task was shown to half the S's and not to the others. All S's performed on the pursuit task for three five-minute continuous practice sessions separated by five-minute intervals. For the film group, the three parts of the film (each part presenting the whole list of instructions) was shown during the five-minute intervals; the control group had "rest" periods during these intervals. S's were scored for the extent to which they employed the techniques depicted in the film as well as for amount of time on target during each successive 30-second "trial" of the pursuit sessions. Twenty student nurses served as S's in the experiment.

It was found that in total on-target time the no-film group's mean was slightly higher than that of the film group (though not significantly). In terms of specific hypothesized detrimental effects, it was found that there was a leveling effect on potentially superior S's at least during the first session ($p < .05$). There was also significant evidence ($p < .05$) in the shape of the practice curve to indicate that the film-showing produced a detrimental massing effect. Anxiety-level ratings were also found to be significantly higher in the film group. The shapes of the learning curve gave no evidence of an hypothesized detrimental effect due to alteration of illumination from viewing to performance conditions.

CHAPTER 13

SOME FACTORS INFLUENCING THE EFFECTIVENESS OF DEMONSTRATIONAL FILMS: REPETITION OF INSTRUCTIONS, SLOW MOTION, DISTRIBUTION OF SHOWINGS, AND EXPLANATORY NARRATION[1]

William J. McGuire

[Editor's Note: The two main sections of this chapter, following the introductory description of design and procedure, deal with the analysis of several factors that were studied in an experiment employing a demonstrational film used to complement student practice on a pursuit-rotor tracking task. Section I analyzes the influence of amount of demonstrational support (number of times the demonstration was repeated) in relation to the use of augmentation of the original demonstration by explanatory narration. Section II reports the investigation of two additional factors: use of slow-motion photography in the demonstration (as compared with normal-speed photography), and spaced showing of the demonstration (with an intervening practice-test session) versus massed showing preceding any practice. The use of an interlocking multi-factor design, and of multiple criterion measures chosen for analysis in relation to the specific hypotheses under consideration, made it possible to obtain the data for these studies in a single, four-replication experiment (which also yielded the data for a further analysis of effects on individual component items of instruction used in Chapter 14). The introductory section of the present chapter describes those aspects of experimental design and procedure which apply in common to the background considerations and analyses subsequently presented with respect to the specific questions dealt with in Section I and Section II.]

Introduction: An Experiment on Factors in the Design of a Demonstrational Film

Film Materials

Twelve directions on how to perform at the pursuit-rotor task were presented to the subjects (S's) by means of an instructional film. The twelve instructional items depicted in the film were the following: (1) keep the back of the hand up; (2) grip the handle with all four fingers and thumb; (3) keep the rod and handle in line;

[1]The writer is indebted to J. J. Wulff, F. D. Sheffield, and G. A. Kimble for advice and suggestions on various phases of the work reported here; see also Footnote 1 to Chapter 12. This chapter was combined by the Editor from two papers provided by the author, which were revisions of earlier Laboratory Notes (McGuire, 1955c,d); these, in turn, were based on drafts originally prepared by McGuire as contractor's reports at Yale University in the summer of 1953.

(4) keep both feet equidistant from the rotor; (5) stand 12 inches from the rotor; (6) stand directly in front of the rotor; (7) stand up straight; (8) use both arm and shoulder; (9) relax; (10) keep moving after the target; (11) follow the path of the target; and (12) hold the wire in your other hand. (The first 11 of these items were employed also in the study reported in Chapter 14.) The revised film used in the present study consisted of four sections, in each of which all 12 of the directions were given. The first time through they were shown in connection with a demonstration by an expert; the second time, by presenting a beginner who was violating most of them; the third and fourth times, by showing this beginner after he had become more proficient and was following the instructions correctly. The film was made in two main parts so that it could be interrupted after Part I (the first two sequences). On all four presentations of the 12 items, the items were given in the same order, for about the same periods of time, and from the same camera angle and distance. (Cf. also Table 1 of Chapter 14.) Each of the four presentations took a total of about two and a half minutes; thus Part I and Part II each required five minutes for showing.

There were two versions of the film. These differed in that, for one of the versions, all the action was shown at regular speed, while in the other version five of the instructional items were depicted in slow motion (about one-third the actual speed of performance). The slow-motion items were (1) "keep the back of the hand up," (2) "grip the handle with all four fingers and thumb," (3) "keep the cord and handle in line," (10) "keep moving after the target," and (11) "follow the path of the target." The other seven items (cf. Chapter 12 or Table 1 of Chapter 14) were shown at standard speed in both versions.

Design

A 2 x 2 x 2 factorial design with an additional control (no-film) group was used for each of four replications. The three factors were: (a) film modality, showing the film depicting the correct performance together with descriptive narration pointing out the correct features in the performance, versus showing the film without the attention-directing aspects of the narration[2]; (b) regular film speed for all items in the demonstration, versus use of slow motion for some of the items (see above); and (c) distribution of showings, showing both Part I and Part II (all four demonstrations) as a single film in one consecutive session, versus distributed showing of the two parts with a 10-minute interval between them. The eight experimental film conditions employed represented all combinations of these three factors. A total of 48 *S*'s were used, four in each of the eight film conditions and 16 in the no-film, control group. The *S*'s were male college students who had volunteered to serve in the experiment in connection with a psychology course in which they were enrolled. A randomized block design was used, with four replications, within each of which one *S* was assigned randomly to each of the eight experimental sub-groups and four *S*'s

[2] Actually, the "film-only" condition did include narration during an introductory sequence, but narration was omitted during the four sequences when the 12 instructional items were demonstrated.

TABLE 1

Experimental Design Used for the Investigations Reported
Sections I and II of This Chapter

(a) Film Modality:		Silent-Film Demonstrations				Film With Narration				Control
(b) Speed:		Regular		Slow		Regular		Slow		
(c) Showing Condition:	(Re Sec I): (Re Sec II):	(2) (Massed)	(1) (Dist.)	(2) (Massed)	(1) (Dist.)	(2) (Massed)	(1) (Dist.)	(2) (Massed)	(1) (Dist.)	(no film) (" ")
(Replica-tions)	1st	7	9	5	8	3	12	6	4	1 2 10 11
	2nd	13	19	15	16	20	14	17	23	18 21 22 24
	3rd	32	34	36	30	28	26	33	25	27 29 31 35
	4th	40	45	42	46	38	44	43	48	37 39 41 47

* Numbers in the cells, randomly assigned by blocks (replications), indicate which S (the number referring to order of participation in the experiment) was assigned to the cell in question.

to the no-film control group.[3] These features of the experimental design are summarized in Table 1, in which it will be noted that the categories for factor (c)—showing condition—are identified separately with respect to the two investigations presented in Sections I and II. The reasons for this dual labeling relate to the differences in purposes and execution of the two analyses, and will be clarified in relation to procedural aspects of the experiment, described below.

Procedure

The sequence of events for each S in all of the film conditions is outlined in Table 2. It will be noted that all S's received two "practice/test" sessions, during

TABLE 2

Sequence of Procedures

Cumulative Time period (approx.)	Film Conditions	
	Massed-Showing Condition	Distributed-Showing Condition
5 minutes	Part I of Film*	Part I of Film*
10 "	Part II of Film**	Rest (5-6 min.)
11 "	Rest (1 min.)	- - -
16 "	First Practice/test Session	First Practice/test Session
21 "	- - -	Part II of Film**
26 "	Rest (10 min.)	Rest (5 min.)
31 "	Scored Practice/test Session	Scored Practice/test Session

* First two repetitions of the demonstration—see text
** Second two repetitions of the demonstration

[3] A randomized-block design was used in order to secure more homogeneous sets of experimental units to which to assign the various treatments. Use of the randomized-block design rather than a completely randomized design was considered desirable to reduce error variance due to variation in subjects and conditions over time, and in particular that due to practice effects, as the experiment continued, among the judges who observed the students' behavior with respect to the various postural-motor instructions given in the film. [Such a design, of course, also has the further important potential advantage of allowing interim analysis following the completion of successive replications, with the possibility of also employing a sequential sampling plan in which the number of replications ultimately run before terminating the experiment can be decided on the basis of the degree of stability (i.e., sensitivity, or "power," of the statistical comparison) attained by the results by the end of any particular replication.--Ed.]

which S practiced while two judges concurrently observed and rated his behavior with respect to the 12 instructional items. For investigation of the effects of amount of demonstration, presented in Section I—two repetitions of the demonstration (totaling five minutes) versus four repetitions (totaling 10 minutes)—it is clear that the applicable data are those from the first practice/test session, since at that point the treatment accorded the groups in the two showing conditions differed only with respect to amount of demonstration that had been given. On the other hand, for the analysis of massed versus distributed showing, presented in Section II, it is evident that the data for the second practice/test session are primarily relevant, since the factor of distribution of showings came into operation only after S's had seen the total film—either in a single, continuous showing (massed condition), or in two parts separated by an interval.

Instructions and Measurements

Each S who had been randomly pre-assigned to any one of the eight film conditions was told he would have a test on a pursuit task after seeing an instructional film about this task. He was shown the film appropriate for his group and then was given a test on a typical pursuit-rotor apparatus described in Chapter 12. As indicated in Table 2, the S's in the "massed" or whole-film condition were shown the entire 10-minute film (Parts I and II) in one continuous showing, and then performed almost immediately in the first of their two five-minute practice/test sessions, with a 10-minute rest before the second such session. The S's in the "distributed" condition were shown the five-minute part of the film; then, after a five- or six-minute rest, they performed in a five-minute practice/test session. (Their performance in this first test session was, as noted above, used to investigate the effect of increased amount of demonstration, at the expense of reduced rest, as described in Section I below.) They were then shown the second half of the film and given a five-minute rest (to make a total of approximately 10 minutes between sessions for this "distributed" condition also); then they performed in their second five-minute test/practice session. (As indicated above, their performance in this second session was used for the analysis of the effect of distributed showing, presented below in Section II.)

Those S's assigned to the control, no-film, condition were simply told:

> "Your job will be to perform on a pursuit task involving this apparatus called the pursuit rotor. You are to hold this stylus (indicate) in your preferred hand and try to keep this point of it in contact with this dime-sized target disk (indicate) as much of the time as possible. The turntable with the target disk will be rotating in a clockwise direction at a constant speed of one revolution per second throughout your test session."

The control S was then given two five-minute tests, separated by a 10-minute rest.

Two types of measures of performance were taken: the pursuit-score (number of seconds of contact) during each 30-second trial (see Chapter 12) of each five-minute session; and the postural-performance score, indicating the extent to which the S adopted each of the 12 postural items presented in the film, as rated by two judges. Judgments on each item were made by each judge twice during each session, the S being given ratings of 1, 1/2, or 0 depending on how well he met the pre-set criteria. The judges had pre-training in evaluating S's conformity to the fairly unambiguous criteria set up for scoring the various items. In addition to the ratings on individual postural items, a total postural score for each S summed over the 12 items was also computed for use in the analysis reported below.

Section I. Effects Produced by Repetition of Instructions Within a Demonstrational Film

Rationale and Background

"Repetition is the mother of retention" is a maxim that might evoke little controversy among students of human learning. Indeed, much of the research in the past half-century that falls under the "learning rubric" has taken for granted the principle that learning is a positive function of repetition, and has concentrated on determining the parameters of this function and the conditions that affect them.

Some exceptions are recognized, of course. For example, it is generally assumed that the performance-practice curve is negatively accelerated, and thus that the practice increments become negligible after a point. (In fact, some current mathematical-model learning theorists seem to be preoccupied with this asymptotic score to the exclusion of the other parameters of the curve, thus largely ignoring the pre-eminently "learning" segment of the learning curve.) There are even special conditions where performance is a negative function of amount of practice; for example, when the "negative-practice" conditions obtain (Peak, 1941), or when work decrement begins to exceed acquisition increment, as is not uncommon during massed practice on a perceptual-motor task (Ammons, 1947b). It is also possible that repetition could, through boredom or the like, produce sufficient hostility in the learner so that the net effect of continuous repetition would be detrimental to learning (Ehrensberger, 1945).

A number of difficulties arise when we attempt to apply laboratory findings regarding the effects of repetition to the use of programmed instructional media. The laboratory findings are often based on data from non-human species and (even in the case of human S's) from the tasks that involved meaningless materials; also they commonly involve performance under motivational conditions quite different from those obtaining under the low-surveillance conditions in mass-instructional situations involving audio-visual aids. It does seem reasonable to expect that in general there would be some learning benefit resulting from repetition of the material within an instructional film, but questions arise, such as: whether the magnitude of the gain is worth the added expense in production costs and audience time involved in lengthening the film; whether it might not be more efficacious to devote any such lengthening to material other than repetitions of material already presented (see Lumsdaine, May, and Hadsell, 1947, 1958); and what other factors might interact with repetition in determining the amount of gain (see Chapter 17).

Some predictions regarding the interactions and limitations of repetitions within the film can be educed from the general experimental literature on learning. First, it is generally found that continued repetition adds successively smaller increments (Gladstone et al., 1947). Hence, continued lengthening of the film would be expected to yield diminishing returns [as has been demonstrated for various kinds of film instruction by McTavish (1949), Lumsdaine, Sulzer, and Kopstein (see Chapter 17), and others]. Secondly, the benefit of additional repetition tends to be greater as the material to be learned is more difficult (Cook, 1937). [This has also been borne out, for film instruction, particularly for the more intelligent learners, in the findings by Lumsdaine, Sulzer, and Kopstein in Chapter 17.] Thirdly, the benefit is increased by distributing the repetitions over a longer time interval rather than showing them in immediate succession [cf. the film-research findings of Miller and Levine, 1952], and this is particularly true with harder tasks (Hovland, 1940). Lastly, the benefit of additional repetitions tends to be greater (particularly for easy material) for long-term retention than for immediate recall (Krueger, 1929). The present analysis provides only a preliminary investigation of the relevance of these findings to the use of instructional media for teaching perceptual-motor skills.

This section is restricted to a test of two hypotheses: (1) That lengthening a film to allow repetition of the instructions has detrimental as well as beneficial effects on performance as compared with simply allowing the subject (S) to rest during the additional time needed for increasing the film length. (2) That the benefits of lengthening the film to allow repetition are less pronounced when the film is accompanied by an instructional narration than when it is a silent film.

As regards the derivation of the first hypothesis, the beneficial effects of repetition within films (or of multiple showings of a single film) have already been amply demonstrated—e.g., Lumsdaine, May and Hadsell (1947, 1958), Jaspen (1949, 1950a), McTavish (1949), and Sulzer, Lumsdaine, and Kopstein (1952). The prediction of concomitant detrimental effects derives from a previous study (see Chapter 12) in which it was found that viewing an instructional film immediately before performance on a perceptual-motor tracking task tended to produce deleterious effects such as are usually associated with massed practice. Specifically, in the practice curve for tracking skill the work decrements were more pronounced and appeared earlier when the interval just before the performance was used to show an instructional film on the task than when the S spent that interval resting (or engaged in task-irrelevant activities).

The second hypothesis—that the benefits of repetition would be greater with a silent film than with a film plus narration—was derived from the assumption that the instructional narration provided in itself a concurrent "repetition" of the visually depicted instructions, and that performance is a negatively accelerated function of repeated practice (whether implicit or overt). The experimental materials, design, and procedure used in the investigation of these two hypotheses have been described in the introductory section of the chapter. It will be recalled that only the measures for the first test session are relevant to the inquiry concerning amount of repetition. Accordingly, in the present section the pertinent aspects of experimental design could be considered, without reference to the variable of slow motion (or of distribution of practice, as such), as a 2 x 2 design in which the four cells represent the four combinations of long (four-repetition) versus short (two-repetition) films, and of narrated versus silent film. Within each of these four cells, data are available for eight S's, in addition to that for the 16 additional S's who were tested in the control (no-film) condition.

Results

The basic data for testing the hypotheses are the posture scores—that is, the judges' ratings as to how well the S's followed the 12 instructional items in the film, as described in the procedure. Some evidence for the reliability of these ratings is indicated by a high correlation (Pearson r) coefficient of +.93 between the sets of total scores by the two judges. The subsidiary score, the pursuit-contact time, did not seem to be a good index of what was taught by the film, since the 48 pursuit-scores correlated only +.27 with the 48 posture-scores (as averaged for the two judges for each S). This lack of substantial correlation is in keeping with the result of the previous study (Chapter 12), in which the correlation was +.12. Hence, it would seem that the hypotheses could be evaluated primarily in terms of the posture-scores.

The posture-score results relevant to the present investigation are summarized in Table 3-A. It can be seen (cf. Table 3-B) that seeing the instructional film did significantly improve performance ($p < .001$) over the control (no-film) group's performance. However, no appreciable increase in effect for the 10-minute film (with four repetitions) was found (Table 3-C) as compared with seeing the five-minute film (with two repetitions). Those who saw the film with descriptive narration did significantly better ($p < .05$) than those who viewed the silent film. Hence, such benefits as are derived from repetition of instructions seem greater here when this "repetition" takes the form of simultaneous, bi-modality presentation than when implemented as successive, one-modality presentations. The interaction of repetition with modality is non-significant.

The predicted detrimental effects of viewing the instructional film just before performance on the tracking task might be more evident in the pursuit scores than in the postural scores. The mean pursuit scores (number of seconds contact with the target during the five-minute trial) are shown in Table 4. None of the differences between the means approach conventional levels of statistical significance, not even the superiority of the groups who saw a film as compared with the no-film control group. (An analysis of variance for pursuit scores for both test sessions is given in Table 8; see also footnote 4, *infra*.)

While small in magnitude, the differences in pursuit scores among the various instructional conditions are suggestive in direction, particularly when compared with the postural scores. The mean pursuit scores are higher after the short film than the long, and after the silent film than after the film with narration. That is, the more instruction immediately preceding performance, the poorer the tracking behavior appears to be.

TABLE 3

Analysis of Posture Scores, First Test Session

A. Mean posture scores obtained under the several instructional conditions

Film Modality:	Film and Narration	Film Only (Silent)	Control (No film)
Four repetitions ("massed")	37.37	32.81	
			20.66
Two repetitions ("spaced")	37.25	33.18	

B. Overall analysis of variance (no-film versus all film conditions)

Source of Variation	Sum of Squares	Df	Mean Squares	F	P
Between film and no-film	2243	1	2243	46.96	.001
Residual error (between film and no-film)	2197	46	47.76		

C. Analysis of Variance within Film Treatments*

Source of Variation	Sum of Squares	Df	Mean Squares	F	P
Modalities (M)	148.78	1	148.78	4.85	<.05
Speed (S)	3.13	1	3.13	<1.00	>.05
Repetition (R) (or Distribution)	0.13	1	0.13	<1.00	>.05
(M) . (S)	120.12	1	120.12	3.92	>.05
(M) . (R)	0.50	1	0.50	<1.00	>.05
(S) . (R)	3.78	1	3.78	<1.00	>.05
(M) . (S) . (R)	47.53	1	47.53	1.55	>.05
Replication	226.28	3	75.43	2.46	>.05
Error	643.97	21	30.67	--	--
Total	1194.22	31	--	--	--

*The sources of variance analyzed include also those involved in the investigation reported in Section II. The factors relevant here are underlined.

TABLE 4

Mean Pursuit Scores (mean number of seconds on target during the first five-minute test trial, for subjects in various instructional conditions)

Film Modality:	Film with Narration	Film Only	Control (No film)
Four repetitions (2-Part, 10-minute film)	39.04	41.55	38.87
Two repetitions (1-Part, 5-minute film)	44.76	45.31	

This pattern of results in part is what we might expect if, as hypothesized, the repetitious instructional presentation in the films increased learning of the techniques advocated, but also increased reactive inhibition with respect to the mediating perceptual responses involved both in viewing and performing at the pursuit task.

Discussion

The above results indicate that, with instructional films of the type used in this study, little will be gained in the amount the audience will learn by increasing the length of the films (repetition or demonstration) and augmenting their informational content by narration. Further, the lack of a significant length-modality interaction effect does not indicate that the lengthening is of more value for a silent film than for a film with narration. The greater expense in production cost and in audience time resulting from a longer film for this kind of task seems unjustified. The teaching effectiveness of a short silent film is, however, improved by the addition of a sound track. This provides simultaneous repetition in a different modality instead of doubling the length of the film, which provides successive repetition in the same modality.

These interpretations are offered within the context of several restrictions which the experimental conditions impose on the generality of these results. First, the type of film involved teaches a simple task, and hence, as indicated in the preliminary remarks, the expected efficacy of increasing the number of repetitions would be minimized. The question remains whether the length of the film (more specifically, the number of repetitions of the information) becomes a more important factor for more difficult tasks. Evidence that such is the case is found in the results reported by Lumsdaine, Sulzer, and Kopstein in Chapter 17, and those obtained in Jaspen's (1949) study, in which repetition did increase learning in the case of a task involving assembly of a breechblock mechanism.

Secondly, in the present study either two or four repetitions of the demonstration were given the S's. It may have been with this simple task that there was a benefit from lengthening the film in the initial stage, but by the end of the second repetition, further benefit from lengthening was negligible. McTavish (1949) found, for example, that two showings of some films produced more learning than one, but that further showings produced progressively smaller increments (cf. also Chapter 17). It might be of interest to repeat the present study with a further shortening of the film so as to present only one repetition of the information in the short-film condition.

Thirdly, results deal with the learning or "immediate recall" scores only. It is quite possible in the light of other findings from learning studies (Krueger, 1929) that different results would have been found if we studied the scores of retention over a long period of time rather than the immediate-recall scores. The repetitions with the longer film may result in an over-learning that does not make for better immediate-recall scores, but does result in better retention over a long interval between showing and performance.

Fourthly, here the repetitions were shown in immediate succession. The hypothesis that additional repetitions or showings are of greater efficacy as they are more distributed in time, which is the indication of previous laboratory studies as discussed above, is susceptible to further test.

Summary

To test several hypotheses about benefits and detriments of repetition of instructions within and across modalities by means of a training film, 48 S's were divided into three groups. One group saw a full version of an instructional film on how to perform at a pursuit task, the film containing four repetitions of the instructions. A second group saw a shortened version with two repetitions. The third group saw no film at all. Half the subjects in each film group saw the film with instructional narration, while the other half saw the film without this narration. The film groups showed significantly more learning than the no-film group, and the narration-film group significantly more than the silent-film group. The double-length film was not found to be significantly more effective than the single-length one, nor was lengthening of the film found to be more important when the film was shown without narration than with it. It was pointed out that further study is needed to determine to what extent these results can be generalized to films involving more difficult materials, to even further shortening of the film, and to effect on prolonged-retention scores as compared with the immediate-recall scores used in this research.

Section II. Slow-Motion Depiction, Distribution of Showings, and Explanatory Narration as Stimulus Factors Influencing the Effects of a Demonstrational Film

Rationale and Background

The results of several studies suggest that films that realistically portray proper techniques for performing at motor tasks tend to be overly complex in stimulus input, with the result that teaching effectiveness is diminished. Excessively complex stimulus input resulting in a rate of information presentation far in excess of the viewer's information-processing capacity could produce a detrimental effect, resulting either in a massed-practice effect, or in inefficient habit-formation because the viewer selects a sample of irrelevant cues for perception.

There is some evidence that viewing an instructional film before performance on a perceptual-motor task does tend to produce a significant amount of mass-practice decrement (see Chapter 12). This massing effect would be particularly pronounced as the film length increased and as the practice immediately followed the film-viewing. To investigate it in the present study, some S's were shown a complete ten-minute film and then immediately performed the task; while others were shown the film in two five-minute parts separated by a rest interval, and performed the task after a further rest interval.

The second potential detriment of overly complex portrayal derives from the limited receptive capacity of the viewer. If we grant, with Miller (1956), that the average rate of information-encoding is about seven bits per second for usual human subjects, then it must be recognized that the informational content (in terms of number of potential sensory discriminations) in a realistic portrayal within a demonstration film far exceeds the viewer's capacity. His capacity would probably be exceeded by a realistic film portrayal even were he to sample seven different bits each second, rather than to react (as is much more likely) in accord with viewing habits that result in a highly redundant sampling of information. Of course, his viewing habits may reflect his past experience with training films and other similar instructional media, so that the selectivity tends to increase the likelihood that he will perceive the proper (intended) aspects. Indeed, as usage of a given instructional medium within the culture increases, so that our population of viewers becomes more familiar with the medium and our production craftsmen more stereotyped in their adherence to the conventions of their art, we can expect viewing habits to become more and more adaptive in perceptual selectivity.

The fact remains that the viewer tends to lose some information when shown an elaborate portrayal of recommended performance technique, and it may well be that a part of this lost information is task-relevant. Hovland, Lumsdaine, and Sheffield (1949, pp. 158-160) have demonstrated that the selective perception of factual film content sometimes results not only in lessened learning from the film, but also even in acquiring erroneous responses, especially by less intelligent viewers.

In recognition of this stimulus-selectivity problem, film makers and researchers have employed various emphasizing techniques to guide attention to the crucial stimuli within films (cf. Chapters 13 and 17). Such indexing devices as color, captions, introductions, narration, moving camera, closeups, and animation techniques are frequently used. It should be noted, however, that such devices themselves add to the stimulus complexity and hence, unless their attention-directing effect is highly conducive to proper learning, may actually lower the effectiveness of a film (Neu 1950). In the present study the use of narration can be considered as an attention-directing variable. It was predicted that the group with the attention-directing narration would be superior in learning those aspects to which attention was specifically directed, but the no-narration group would be superior in learning the aspects not specifically cited in the narration.

A third variable in the present analysis (in addition to distribution of showing and attention-directing narration) was slow-motion versus normal-speed portrayal. Either of the two postulated detrimental effects of stimulus over-complexity would lead to a prediction of benefits from slow-motion portrayal. Slow-motion portrayal should reduce the rate of information presentation and hence produce a distributed-practice effect. This should result in better learning of the instructions portrayed in slow motion and also, though to a lesser extent, of the instructions portrayed at regular speed as well. Also the pursuit-scores after the slow-motion showing should approximate the typical distributed-practice curve (characterized by a greater warm-up effect and longer intra-sessional time to reach maximum performance) than do the practice curves of the regular-speed film group (Ammons, 1947a,b). Alternatively, the slow-motion portrayal might serve as an attention-directing device and so result in better learning of the instructions so portrayed, but in no effect or even in a negative effect on the learning of the instructions portrayed at the more typical regular speed. Hence this third variable of speed allows some possibility of detecting which of the two postulated detrimental effects of over-stimulation has the greater impact on performance.

Methods used in investigating these and other hypotheses have already been described. For purposes of the present inquiry the same groups compared on the first test session with respect to film repetition served, on the second test session, to provide the data for analysis of distributed versus massed showings.

Results

Effect of Massed Versus Distributed Showing. The overall analysis for both test sessions for pursuit scores shown in Table 5, and a similar analysis for posture scores, fails to demonstrate a significant difference between massed and distributed groups in either posture or pursuit scores. However, effects of distributed showing could be obscured by the fact that in the experimental procedure some of the performance of the distributed group came after the S's had seen only half the film, while all the performance of the massed group came after seeing the whole film. To allow for this possibility, relevant mean pursuit and posture scores were computed (see data in the top parts of Tables 6 and 7) for the second session alone, since all the S's, whether in the massed or distributed conditions, had seen both parts of the film by the time they performed in the second session. Some indication of superiority for the distributed group over the massed group in the second-session pursuit scores is suggested by the means in the top part of Table 6; and a similar indication can be seen for the posture scores from the corresponding means in Table 7. However, the F of only 1.23 for the latter (as shown in Table 8) falls short of conventional significance levels.

It will be noted that the difference with respect to pursuit scores between the massed- and distributed-practice groups is in the same direction as that with respect to posture scores. As can be seen in Table 6, the second-session difference between

TABLE 5

Analysis of Variance for the Pursuit Scores
(Both Test Sessions)

A. Overall Analysis

Source of Variation	Sum of Squares	Df	Mean Squares	F	P
All treatments vs control	1254.26	1	1254.26	0.44	>.05
Error	131677.19	46	2862.55	--	--
Totals	132931.45	47	--	--	--

B. Within Treatment Analysis

Source of Variation	Sum of Squares	Df	Mean Squares	F	P
1. Modalities (M)	17.21	1	17.21	0.02	>.05
2. Speed (S)	2393.54	1	2393.54	2.65	>.05
3. Distribution (D)	1196.38	1	1196.38	1.32	>.05
4. (M) . (S)	278.59	1	278.59	0.31	>.05
5. (M) . (D)	212.06	1	212.06	0.23	>.05
6. (S) . (D)	56.98	1	56.98	0.06	>.05
7. (M) . (S) . (D)	2311.80	1	2311.80	2.56	>.05
8. Replications	12061.18	3	4020.39	4.45	--
Error for 1 to 8	18961.71	21	902.94	--	--
9. Sessions (T)	69001.43	1	69001.43	122.31	.001
10. (T) . (M)	105.96	1	105.96	0.19	>.05
11. (T) . (S)	85.72	1	85.72	0.15	>.05
12. (T) . (D)	244.10	1	244.10	0.43	>.05
13. (T) . (M) . (S)	9.64	1	9.64	0.02	>.05
14. (T) . (M) . (D)	340.54	1	340.54	0.60	>.05
15. (T) . (S) . (D)	48.35	1	48.35	0.09	>.05
16. (T) . (M) . (S) . (D)	18.55	1	18.55	0.03	>.05
Error for 9 to 16	13539.19	24	564.13	--	--
Totals	120882.93	63	--	--	--

TABLE 6

Mean Pursuit Scores Obtained in the Second Test Session by S's in Each Sub-Group for Each Experimental Condition, and During Both Sessions (totaling ten minutes of practice) for Relevant Comparisons

Test Session	Spacing	Modality and Film-Speed of Demonstration: Film + Narration Reg.	Film + Narration Slow	Film Only Reg.	Film Only Slow	Control (no-film)
Second Session	Massed	89.87	126.09	96.08	96.19	---
	Distributed	110.91	113.68	107.39	126.46	---
	Av. M & S	100.39	119.88	101.24	111.32	101.28
First Session	Av. M & S	35.24	48.55	40.17	46.69	38.87
Av. of Both Sessions	Av. M & S	67.82	84.22	70.70	79.01	70.08
		76.01		74.86		

TABLE 7

Mean Posture Scores Obtained by S's in Various Experimental Conditions (Scores in the cells are mean ratings of adherence to technique advocated in the film.)

Test Session	Distribution of Showings	Modality and Film-Speed of Demonstration: Film with Narration Reg.	Film with Narration Slow	"Silent" Film Reg.	"Silent" Film Slow	Control (no-film)
Second Session	Massed	38.38	39.38	37.62	34.25	---
	Distributed	42.38	38.88	39.25	38.12	---
	Av. M & S	40.38	39.13	38.44	36.18	23.25
First Session	Av. M & S	35.16	39.56	34.62	31.37	20.66
Av. of Both Sessions	Av. M & S	37.72	39.34	36.53	33.77	21.96
		38.53		35.15		

TABLE 8

Within-Treatment Analysis of Variance for the Posture Scores, Second Test Session Only

Source of Variation	Sum of Squares	Df	Mean Squares	F	P
Modality (M)	47.53	1	47.53	1.45	>.05
Speed (S)	24.50	1	24.50	<1.00	>.05
Distribution (D)	40.50	1	40.50	1.23	>.05
All Interactions	29.32	4	7.33	<1.00	>.05
Replication	99.16	3	33.05	1.01	>.05
Error	686.46	21	32.69	--	--
Totals	927.47	31	--	--	--

the two groups is substantial (the distributed group averaged 13 per cent more contact time than the massed group). However, due to the high inter-S variance at this pursuit task, the significance of this difference is trivial ($F < 1.00$). The difference is more pronounced during the first half of the second session, which we would expect on theoretical grounds. During this half-session, the distributed-film group's superiority to the no-film group attains borderline significance ($t=1.90$), while the massed-film group's superiority to the no-film group is trivial ($F < 1.00$). However, even during this half-session, where the difference is most pronounced, the superiority of the distributed over the massed group attains only an unimpressive level of significance ($t=1.24$). The F of the difference between pursuit scores of regular-speed and slow-speed film groups in the second session only was 1.36. That the pursuit scores were so insensitive to the experimental manipulations is not surprising in view of the low correlation ($r=0.27$) between the posture scores and these pursuit scores, and in view of the high error (individual difference) variance with the pursuit task. However, the differentials due to the several film variables in pursuit scores, though not significant in magnitude, all agree in direction with the more significant differentials in posture scores.

Effect of the Narration Variable. Narrations were actually included in both versions of the film, but their instructional explicitness differed. The no-narration film provided a narration that simply announced that the film provided instructions for performing at the pursuit-rotor task. Then, before the beginner's performance was depicted in the film, it was further announced that a subject would be shown performing poorly and, in two subsequent sessions, that he would be shown performing correctly. It is called the no-narration version since none of the 12 specific instructional items was explicitly designated in the narration. The narration version, on the other hand, called the S's attention to 11 of the 12 crucial aspects of the depicted performance. Since the no-narration version did have some auditory accompaniment, the finding of a significant difference between it and the full-narration version is perhaps the more persuasive.

Comparison of Tables 7 and 8 with Table 3 shows that the mean posture scores were higher for the narration group than for the no-narration group in both practice sessions, but that this difference was significant ($p<.05$) only for the first-session analysis. (P was also $<.08$ for the combined-analysis of posture scores for both sessions.) The superiority (at least in the first session) attributable to the added narration indicates that, for the type of film used in the present study, the addition of narration can increase the instructional value of the film even though it only calls attention to what is being depicted on the screen.

One might expect that the instructional increment supplied by narration would be concentrated in those items of the instruction that are hard to demonstrate visually. To test this hypothesis, an item-analysis was done to find which items benefited most from the addition of the narration. A sign test of the differences between the two conditions shows that the added narration causes a significant (.05 P-level) improvement in postural Items 1, 4, and 6. An independent judgment of the film indicates that Items 1 and 6 were the poorest of the twelve items in visual depiction, tending to confirm the hypothesis, though Item 4 does seem adequately portrayed in several scenes.

This finding of the superiority of film and narration over film may be interpreted as confirming the finding of Nelson and Moll (1950), who found dual-modality presentation to result in higher scores on an information test dealing with contents of films on flight theory and on survival in the desert. The present study goes beyond this previous work in showing that the advantage is found not merely in an information-recall test, but rather in scores of how well the audience actually puts that information into practice, i.e., adopts the recommended postural techniques. The present study also extends the finding to another type of film, one teaching a perceptual-motor task.

Perhaps the most interesting finding in connection with the item-analysis between the film-only and the film-and-narration groups has to do with the effect on Item 12. This item differed from the other eleven in that it was clearly depicted visually, but was not mentioned in the narration. It might be thought that since it was not mentioned in the narration the film-and-narration S's would score no higher on this item than would the film-only group. In fact, however, the results indicate that for this item the film-and-narration S's actually did worse than the film-only S's, while the combined mean for the other items is decidedly in the opposite direction. The analysis shown in Table 9 indicates that this difference is significant beyond the .05 level (as manifested in interaction effect between the film-only versus film-and-narration condition and the Item 1 to 11 mean versus Item 12). Such a result tends to confirm the hypothesis that any information conveyed in the film is more likely to be derived by the S from a silent film than from a film with narration when the narration does not explicitly point up this particular item of information. The narration serves as an attention-directing device for the items explicitly mentioned, but at the expense of the depicted but unmentioned remainder of items. Hence it is important, if narration is to be used with the film, that all the important information be conveyed in the narration as well as visually. Otherwise, while the information presented in both auditory and visual modalities will be better grasped with the added narration, the information conveyed only visually will actually be less well learned from a film with narration than it would have been from a silent film.

TABLE 9

Analysis of the Difference between Item 12 and the Other Items and Its Interaction with the Audio-Visual Versus Visual-Only Modalities

Source of Variation	Sum of Squares	Df	Mean Squares	F	P
Modality (film only versus film and narration)	42.00	1	42.00	10.91	< .01
Item (composite 1 to 11 versus 12)	614.41	1	614.41	159.59	< .001
Interaction: modality with item	16.64	1	16.64	4.32	< .05
Error (S's within condition for the item)	1464.05	380	3.85		
Total	2137.10	383			

Effect of Slow-Motion Versus Regular Speed. The overall analyses (cf. Tables 4 and 8) fail to establish significantly better total postural scores in either session for the group which saw parts of the film in slow motion over the group which was shown all action at regular speed. A more precise estimate of the effect of this variable is available in that all S's in the slow-motion condition were shown some instructional items in slow motion and some at regular speed, and hence served "as their own control" with respect to this variable. Hence, a sensitive test would be a "mean-difference" test between the slow speed items versus the regular speed items within each S in the slow-motion treatment. Since the items were not rotated between these conditions, there is the possibility that any such difference would be due to difference in the initial difficulty of one set of items rather than to the special difference between the sets. To control for this possibility, a second order difference-between-differences test was made, using the difference between the two sets of items in the group which saw the whole film at regular speed (to eliminate the effects of any difference in initial difficulty between the two sets). This analysis showed that there was a significantly (.01 P-level) greater superiority of learning of these critical items when they were depicted in slow motion than when they were shown at regular speed.

Originally, it had been hypothesized that depicting some of the performance in slow motion, thus slowing the rate of information presentation, could have a beneficial effect in either of two ways, by distributing practice or by calling the S's attention to relevant cues. It was suggested that the relative importance of these two hypothesized mechanisms might be assessed by examining how showing some of the items in slow motion affected the learning of the other items shown at regular speed, even in the slow-motion occasion. Insofar as distributing practice was involved, learning of these other items should, if anything, have been augmented; insofar as attention-directing was involved, learning of these others should, if anything, have been impeded. The results tend to show the latter effect. The seven items shown at regular speed in the slow speed version were more poorly learned in that version (.10 P-level, two tails) than the same items in the version in regular speed throughout. Here, as was also the case with the added narration, we find the attention-directing device within the film not adding to its overall teaching effectiveness, but rather causing some content to be better learned at the expense of the remainder.

Further doubt is cast on the other hypothetical effect of slow-motion showing —that it distributes practice—by the failure to find any evidence, in the practice curves for the pursuit-contact scores, for greater reactive inhibition with the regular speed than with the slow-speed showing. There was no evidence of the predicted (Ammons, 1947a) greater warm-up increment or greater intra-sessional time to reach maximum performance in the slow motion than in the regular-speed showing.

Discussion

The evidence regarding the operation of one possible detrimental effect of excessive stimulus input via the film—namely, that it caused undesirable massing of practice—was largely negative. Neither the distributed showing of the whole film nor the use of slow-motion sequences within the film produced any appreciable effect that could be ascribed to distribution of practice. An earlier study by Ash (1949) also resulted in negative findings regarding the difference between massed and distributed film showings. Ash's study involved the effect on information scores of showing a number of different films massed or with intervening intervals. The efficacy of distributing the material to be learned has been found to be directly proportional to the homogeneity of the material (Hovland, 1938). Therefore it might be expected that the present study would be more likely than the Ash study to show an advantage in the distributed-showing condition, since in the present case it was the quite homogeneous contents of a single film that were massed or distributed, while in Ash's study it was several different films, much more heterogeneous in subject matter, that were massed or spaced. Hence in the present case, conditions of film-showing and subsequent performance were extremely favorable for the appearance of an advantage in the distributed condition, and the negative results strongly suggest that this variable may be a relatively unimportant one in the use of instructional films.[4]

[4] [Cf. also, however, the positive results on spacing or "distribution" factors in film instruction obtained by Miller and Levine (1952) and by Faison, Rose, and Podell (1955); see abstracts of these studies in the Appendix. — Ed.]

There is, however, a consideration in connection with the present study which might call into question the above evaluation of the distribution-of-showing variable. The learning from the film (in terms of posture scores) appears to have reached an asymptote by the end of the first half of the film (cf. Table 3). Hence, it might be said that the reason the spaced showing of the second half failed to show a significant advantage over the immediate showing was that little learning occurred from the second half in either case. To test this alternative interpretation and further maximize the possibility of an appearance of a favorable effect due to distribution, the spacing interval should occur at a point in the film beyond which a considerable amount of further learning has been found to take place. Alternatively, instead of an immediate test, a test of retention over a considerable period might be made. In this manner, also, the manifestation of any possible advantage in spacing the presentation would be favored, since—even if the part of the film after the interval does not add any immediately measurable new learning—it should cause "overlearning" of the items already learned, and such overlearning has been found to result in superior long-time retention (Krueger 1929, 1930b). All of these suggestions in the way of favoring the appearance of any effect due to distribution of showing are made because, in view of the logical impossibility of proving the null hypothesis—viz., that there is no advantage to distributed showing—the possibility should be rejected in theory only after every possibility of its manifesting an effect has been provided for. We would not expect any very sizable effect of distribution since the covert, vicarious practice involved during film-viewing generally does not have great response similarity to the overt practice behavior. Nevertheless, these covert responses presumably play some mediating role in overt performance (if we are to account for the often-found efficacy of instructional film-viewing on subsequent motor-skills performance), and hence any reactive inhibition developed in connection with them would have some detrimental effect on overt performance.

The present results indicate, as do those of earlier studies (cf. Chapter 17; also Lumsdaine, May, and Hadsell [1947], and Ch. 7 of May and Lumsdaine, [1958]), that appropriate attention-directing devices can increase the learning of the items to which attention is called. The increased learning produced by added narration and slow-motion portrayal was found to be at the expense of learning items to which attention was not called. The result fits the theory that the viewer's information-encoding capacity is far below the rate at which information is presented by a realistic training film. Hence such attention-directing devices merely focus learning on part of the content at the expense of the rest.

In fact, the device itself adds to the complexity of the film and may therefore result in poorer learning, not only of the remainder of the items, but also even in less total learning, since part of the S's perception is diverted to the device itself. To test for this more serious detriment we need to compare total learning, for stressed items and other items combined, in the film version using these attention-directing devices and the version not using them. A further requirement is that the amount of film instruction given on the two sets of items be equal. A test was not possible with the present data because this latter condition was not met.

Even though we cannot conclude that the net effect of attention-directing devices on content learning is negative, the evidence is clear that their effect on the items to which attention is not called is negative. Hence these devices are useful only to the extent that those items to which attention is directed are more important for good performance than are the items to which attention is not called. One of the unique advantages of training films in depicting performance realistically is that they allow us to instruct S by showing him good (successful) and poor performance even when we do not know precisely the specific characteristics which make these performances good or poor. (Analogous here is the case-history method in law, medicine, or business administration to communicate principles that the instructors themselves have not tested out explicitly.) It is with this type of pedagogic problem that the use of attention-directing devices is particularly likely to be detrimental.

Summary

A postulate that realistic film portrayals of performance on motor-skills tasks tend to present information at a rate in excess of the S's receptive capacity led to several predictions about distribution of showing, the use of an attention-directing narration, and slow-motion portrayal of some of the performance. Two possible mechanisms in the hypothesized effect of these variables is that they influenced development of reactive inhibition or that they resulted in selective perception of the relevant content.

Distribution, narration, and slow motion were each employed as two-level film variables in a 2 x 2 x 2 factorial design with an added control (no-film) condition. The films taught some recommended techniques for performing at the pursuit-rotor task. After viewing the film, the S's performed at the task, while being scored for utilization of the recommended techniques and also for pursuit-contact scores.

It was found that the film did affect the mode of performance ($P < .001$). Distributed showing produced, overall, little superiority to the massed showing. The attention-directing narration improved learning of the items to which attention was directed, but impeded learning of the other items. Likewise, depicting some material in slow motion and the rest at regular speed improved learning of the former and reduced learning of the latter. These results were interpreted as indicating that the mechanism that made the several cinematic devices effective was their attention-directing rather than practice-distributing effect.

CHAPTER 14

EFFECTS OF SERIAL POSITION AND PROXIMITY TO "REWARD" WITHIN A DEMONSTRATIONAL FILM[1]

William J. McGuire

Introduction

Two variables that have been found to be related to learning by rats and humans under simple, well-controlled laboratory conditions are investigated here for their relevance to more realistic learning situations involving audio-visual training aids. These learning variables are (1) nearness to some "reward" sequence within the film, and (2) serial position within the film.

The present analysis must be regarded as exploratory, since it employed material far from ideal for the above purposes. Perhaps the most serious limitation is that, for practical reasons, we were not able to construct the many different versions of the film that would have been necessary to counterbalance all the materials among the treatments. Hence it was necessary to control for some extraneous factors by the less satisfactory procedure of *post-factum* statistical adjustments.

Method

Materials

A ten-minute film was used to present a series of instructions on how one should perform at a perceptual-motor task involving keeping contact between a stylus held in the hand and a target disk located on a rapidly rotating turntable. The specific apparatus shown in the film and used in the test was a conventional pursuit rotor, with a dime-sized target disk, rotating clockwise at a rate of one revolution per second. (See also Chapter 13.) There were 11 instructional items in all, the entire series being presented visually and in the narration four times during the film. In the first presentation, the series of instructions was shown in

[1]Valuable contributions were also made at various stages of this study by Mark A. May, Fred D. Sheffield, Gregory A. Kimble, J. Jepson Wulff, and Doris Kraeling; see also Footnote 1 to Chapter 12. The present chapter is the author's revision of a mimeographed Laboratory Note (McGuire, 1955e) which was based on the original contract report prepared at Yale University in 1953.

a demonstration by an expert. In the second presentation, they were shown being violated by a beginner at the task. In the third and fourth presentations, the series of instructions were shown being correctly followed by the same beginner after he had become proficient at the task. Table 1 lists the instructional items together

TABLE 1

Instructional Items: Serial Position, Content, Proximity to Reward, Cinematic Characteristics, and Time

Serial Position	Content	Steps Removed From "Reward"	Camera Distance	Total Time Alloted (Seconds)
1	Keep the back of the hand up	3	CU	50
2	Grip the handle with all four fingers and thumb	2	CU	50
3	Keep the rod and handle on line	1	CU	50
	"reward" sequence			
4	Keep both feet equidistant from the rotor	4	CU	25
5	Stand 12 inches from the rotor	3	LS	25
6	Stand directly in front of the rotor	2	LS	25
7	Stand up straight	1	LS	25
	"reward" sequence			
8	Use both arm and shoulder	4	LS	25
9	Relax	3	LS	25
10	Keep moving after the target	2	CU	25
11	Follow the path of the target	1	CU	25
	"reward" sequence			

with their cinematic characteristics in this film. It will be noted that the items were given in the same order in each of the four presentations, and that each item retained constant cinematic characteristics over the four presentations.

On each of the last three presentations, the film performer's score was shown after the third, seventh, and eleventh items. This was done by showing contact time registering on a clock at these points in the film, and stating, during the second (poor performance) presentation, that the performer's failure to follow the instructions was netting him a poor score, and, on the last two (good performance) presentations of the series, that following the instructions was getting him a good score.

Subjects and Design

Thirty-two male college students (who volunteered to participate in the study in connection with a psychology course they were taking) were used in this study. These S's were divided into two groups of 16 each—a control (no-film) group and an experimental group which saw the film. S's in both groups were scored as they themselves subsequently engaged in the pursuit task during two five-minute performance sessions, between which there was a 10-minute rest interval. The scores taken were ratings by two trained judges on the extent to which the S performed the 11 items in the manner taught in the film. As in the study reported in Chapter 13, each item was rated twice, independently by each judge, in each of the two performance sessions. Hence, each S was judged eight times on each item during the experimental sessions. Each time the S was rated 0, 1/2, or 1 for no compliance, partial compliance, and full compliance, respectively, with the instructions. (The pursuit scores were recorded but are not relevant to the hypothesis under consideration and hence are not discussed further.)

Procedure

S's in the control group were given the same instructions as in Chapter 13 and were told that they would be scored on the basis of amount of time they succeeded in keeping the stylus in contact with the disk.

The S's in the experimental (film) condition received similar instructions and, in addition, were told that they would be shown a film presenting the best way of performing a pursuit task. They performed the task after seeing the 10-minute film.

Results

Item Adjustments to Determine Treatment Effects

The postural scores obtained are shown in Table 2, which gives the total number of points obtained on each item. Since each S received zero to eight points on each item, and there were 16 S's in each condition, the maximum possible score on any one item in either group is 128 points.

TABLE 2

Total Postural Score on Each Item Made by S's in the Control and Film Groups, Together with the Predicted Scores for the Experimental Group and the Deviations Therefrom

Items	1	2	3	4	5	6	7	8	9	10	11
Control (no film) N=16	68.0	44.5	48.0	33.0	61.0	50.5	49.5	86.0	77.5	86.5	90.0
Condition Experimental (film) N=16	116.5	118.5	116.5	126.0	105.0	107.5	118.5	104.5	94.5	103.5	112.5
Predicted Experimental Scores*	109.68	117.28	116.14	120.99	111.95	115.34	115.66	103.87	106.62	103.71	102.57
Deviation of Obtained from Predicted Experimental Scores	6.82	1.22	0.36	5.01	-6.95	-7.84	2.84	0.63	-12.12	0.21	9.93

*(The predicted experimental scores are based on overall correlation of -.69 between experimental and control group scores by item.)

The present experimental design was far from ideal for determining the treatment effects, because the items were not rotated over the several experimental conditions involving serial position and proximity to reward. Systematic counter-balancing of the 11 items would have required constructing 11 different instructional films, which was ruled out by practical considerations at this exploratory level of investigation. Hence, the scores in the film condition do not, by themselves, indicate whether the differences obtained are to be attributed to the experimental treatments or to fortuitous differences in inherent difficulty of the items.

However, the scores in the control (no-film) condition allow for post-factum statistical control, which partially compensates for the lack of experimental control of item difficulty. The choice of the precise statistical adjustment to be made is somewhat arbitrary. Simple differences in scores between control (no-film) and film conditions are obviously inadequate, since there is a wide range of differences among items in the control group, and hence we would be dealing with all the problems inherent in comparing changes from a shifting baseline. A somewhat better adjustment would be the "effectiveness-index" (Hovland, Lumsdaine, and Sheffield, 1949), which would correct for the different ceiling effects. Even this index, however, would leave uncorrected a number of other scaling artifacts. It is likely, for example, that the judges consciously or unconsciously shifted the stringency of their scoring criteria from item to item, perhaps by scoring very strictly on items on which even the control group received high scores, and leniently on items on which the control group did poorly.

It was presumed that the obscuring effects of item inequality and the various scaling artifacts would best be eliminated by an analysis of co-variance that would adjust the differences among items in the experimental condition in terms of the differences among items in the control condition and of the overall correlation between the control and experimental scores on the items. Ideally, the correlation should have been obtained from within-group S scores, but this was not possible since control and experimental scores were obtained from different groups of S's, and the S's in these groups were not paired off in any meaningful way. Hence, the correlation was computed on the basis of just 11 pairs of scores—the control and experimental scores on each of the 11 items. The Pearson r was -0.69, the negative direction possibly indicating a shifting stringency by the judges on the basis of the control-group scores. An expected experimental-group score for each item was obtained on the basis of the control-group score on that item and the -.69 correlation. The total variance was computed by calculating the deviations of experimental S scores from the expected score separately for each item, and then summing the variance for the 11 items. In determining the error terms, 11 degrees of freedom were assigned to the calculation of the predicted mean, leaving 154 df's for the error term. The subsequent discussion of treatment effects is based on these adjusted scores.

Effect of Serial Position

To test for the effect on learning of serial position within the film, it is necessary to eliminate the proximity-of-reward effect by grouping the items before each successive reward into blocks. On the basis of the three points so obtained, we find the typical non-monotonic serial-position curve, with the middle items most poorly learned, and the terminal items more poorly learned than the initial ones. In terms of the adjusted deviation scores discussed above, the block of items coming first produced a learning score of 8.40; the middle block, -6.94; and the final block, -1.77. This serial-position effect is significant at the .05 level (Table 3).[2]

TABLE 3

Deviation Scores of Mean Obtained from Mean Predicted Scores for Experimental Group by Item, Showing their Treatment Level on the Two Variables and the Analysis of Variance

Steps Back From Reward	Ordinal Position Series			
	First	Second	Third	Total
4		5.01	0.63	5.64
3	6.82	-6.95	-12.12	-12.25
2	1.22	-7.84	-0.21	-6.83
1	0.36	2.84	9.93	13.13
Totals	8.40	-6.94	-1.77	-0.30

Analysis of Variance:

Source of variation	Sum of Squares	df	Mean Square	F	(two-tails) p
Serial position	36.31	2	18.16	3.43	.05
Distance from reward	138.85	3	46.28	8.73	.001
(position). (distance)	261.71	5	52.34	9.88	.001
Error	816.28	154	5.30		

Effect of Proximity to Reward

The position of the item relative to the symbolic reward within the film also has a significant effect on learning, in fact, a greater effect than did serial position

[2] The serial-position curve can also be detected across all 11 items, though in terms of the more molecular analysis the effect is somewhat obscured by individual item variance and the proximity-to-reward variable. A parabola of "best fit" was computed (by least-squares method) to all 11 points. The equation of this curve is $Y = .27x^2 - 3.61 + 9.33$. The present data obviously lack the precision necessary to generalize the parameters obtained with any confidence.

(Table 3). However, it should be noted that this variable did not produce a simple delay of reinforcement effect (Hull, 1951, pp. 52ff.), but rather something more like the "spread of effect" described by Thorndike (1933), in that the introduction of each of the symbolic reinforcements within the film augmented the learning, not just of the preceding item but of the following item as well. As will be discussed below, there is reason to interpret this variable as manipulating distribution of practice rather than reward.

Discussion

The applicability of the present results is somewhat restricted by the rather tortuous statistical adjustment required to compensate for the lack of counterbalancing of the items (due to practical limitations) within the experimental treatments. For example, use of the cruder "effectiveness-index," discussed above, would have changed the finding regarding the serial-position effect: that index yields a "primacy effect" such that the first block of items shows better learning than the second with the third block showing poorest learning, rather than the more typical non-monotonic curve yielded by the co-variance adjustment. This alternative effectiveness-index adjustment would not, however, have changed the finding regarding the effect on learning of proximity-to-reward within the film.

An additional source of variance in the present study that may have confounded the results somewhat involves several variations in cinematic characteristics of the presentation. There was some difference between items with respect to amount of time they were discussed in the film, the use of close-up versus distance shots, and the number of instructions given within a single sequence without changing camera distance or angle (see Table 1). These factors might possibly have an effect on learning and should ideally be controlled in the research film. A film so mechanically constructed would, however, be likely to have a dampening effect on the S's interest.

There are two reasons for reinterpreting the effect on learning of proximity-to-vicarious-reward within the film as a spacing rather than a reinforcement variable. In the first place, the benefit of these vicarious-reward sequences extends forward to the subsequent items, as well as backward to the preceding items. Thus, within each block of items (between the successive reward sequences) we see the typical low-shaped serial-position effect (Table 3). Secondly, the two major experimental variables—serial position and proximity-to-reward—show a significant interaction effect (Table 3), such as would be expected if the introduction of the vicarious-reward sequence were serving to distribute practice and hence mitigating the interference effects, especially toward the middle of the list where such interference is greatest. On the other hand, even if the efficacy of the vicarious-reward sequences derives mainly from its practice-distributing effect, there is some suggestion of an additional, specifically reinforcement effect in that the retroactive benefits to the preceding items is greater than the proactive benefits to the subsequent items. Another study is needed to compare the effect of introducing vicarious-reward sequences with the effect of introducing alternative non-reward material, which would distribute the presentation of the instructional material.

Summary

Hypotheses regarding the effects on learning to carry out instructions presented in a training film of two within-film variables were tested: serial-position, and proximity to interspersed vicarious-reward sequences. A film was constructed to present four repetitions of a series of 11 successive instructions. During each repetition a vicarious reward (of seeing the actor obtain a good score for carrying out the instructions, or a bad one for not) was presented at three points in the series. S's were 32 college students, 16 serving in the film condition and 16 in a no-film control. The scores in the no-film condition provided a basis for a statistical adjustment in the film group's scores so as to determine the effect of the film itself on the separate items of instruction.

It was found on the basis of the adjusted scores that there was a typical non-monotonic serial-position curve, with the instructions presented near the beginning of the sequences best learned; those near the end next best; and those near the middle most poorly. Proximity-to-reward also contributed a significant amount of variance but this proximity benefited the subsequent as well as preceding items of instructions, thus suggesting that the efficacy of vicarious-reward sequences within the film operates at least in part as a spacing rather than a reinforcement variable.

CHAPTER 15

"RESPONSE GUIDANCE" AS A FACTOR IN THE VALUE OF STUDENT PARTICIPATION IN FILM INSTRUCTION[1]

Gregory A. Kimble
J. Jepson Wulff

Problem

It is clear that the amount learned from films and audio-visual materials can sometimes be increased markedly by the incorporation of "audience-participation" procedures in which the trainee is given an opportunity to practice what is being taught during the film presentation. But though such "learning-by-doing" procedures have sometimes been found to result in marked gains, in other instances the expected advantage has not materialized. These facts seem to make it quite plausible that some forms of student participation are superior to others.

An important possibility is that participation is more effective if it is guided, so that the trainee's participation responses are restricted to the right ones or nearly the right ones. Or the opposite may be true; that is, the better form of participation may be one in which the learner has to figure out the right answer for himself.

This is a report of an experimental comparison of the training effectiveness of guided and unguided participation. Thus, the question asked is this: When trainees are instructed with a film incorporating audience-participation procedures, do they learn more if they are given guidance to foster correct practice, or is it better if the trainee is left to try to achieve the correct performance without such guidance?

[1]The experiment reported in this chapter was under Contract AF 33(038)-13678 between Yale University and the United States Air Force. This contract work, as well as that reported in Chapter 16, was monitored by A. A. Lumsdaine for the Human Factors Operations Research Laboratory of the Air Research and Development Command. This report is a revision of HROFL Memo Report No. 36 (Kimble and Wulff, 1953b), which included an appendix giving detailed supporting tabulations. Kimble is now at Duke University. See also Footnote 1 to Chapter 10.

Method

Experimental Materials

This study investigated the comparative influence of two kinds of participation procedure on a single task—learning to read the "C" and "D" scales of the slide rule. The investigation made use of two specially prepared sets of filmed training material with recorded narration. Both films provided for frequent audience participation in the form of practice exercises. These exercises were interspersed between sections of film instruction, so that after the presentation of each main point by the film, the students had to do an exercise utilizing the information just presented. Each exercise posed a problem and required students to mark their answers in a practice booklet.

In one presentation (Technique A), aids to the performance of correct participation responses were used. In the second presentation (Technique B) no such aids were used, and the subject had to try to make the correct responses unaided. In all other respects the two film presentations were identical; they used the same 32-frame sequence and the same basic narration. (The additional sentences required to call attention to the various hints to the location of the correct responses in Technique A were omitted from the narration for Technique B.)

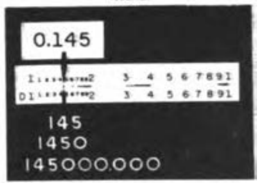
The subjects (Air Force basic trainees) participated by locating numbers on photo-offset copies of the slide rule made up into practice booklets. The booklets for use with slide-film A differed from those for slide-film B only in that the former provided guidance designed to foster correct practice by the learner. These "prompts" were omitted in the practice booklet used with slide-film B. Guidance, or prompting, was given either by indicating the limits of the general region of the scale within which the correct answer to a problem would fall, or by marking and labeling a numerical value related to the correct answer.

The difference between the two procedures is shown in Figure 1 by a series of frames from film A and an accompanying excerpt from the practice booklet. The examples illustrated came at the point in the film instruction after the method of locating integral numbers from one to 1,000 had been explained and practiced. The section illustrated explained how to utilize the information already presented so as to locate numbers involving decimals and numbers larger than 1,000.

Subjects and Training Procedure

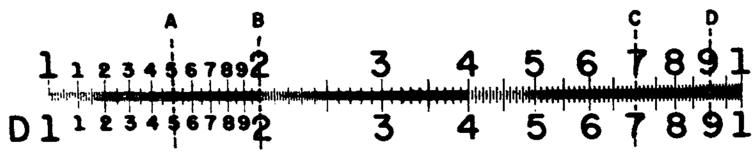
It may be inferred from the description of experimental materials that the experimental design was basically a two-group design. One group was trained using Technique A, the other group using Technique B. Approximately 330 men were trained in each group. The subjects were airmen undergoing classification testing at a basic-training center. The men were drawn from 48 flights of airmen. Only one-fourth of the men in each flight were used in the experiment. Twelve classes, each made up of four independent quarter-flights, were assigned to the experiment. Six of these twelve classes were assigned without bias to treatment A, VISUAL

Narration



By now you have probably noticed that numbers like 145, 1450, and 145,000,000 are all located in exactly the same place on the slide rule.... We would also locate .145 in the very same place since we must use the same rule to locate it.....

On your own scale on page 12 locate .151 (point one-five-one). The correct answer will fall somewhere between "A" and "B" on your scale.



(PAGE 12 OF PRACTICE BOOKLET FOR GROUP A)

As another example, locate .82 (point eight-two) on the same scale. It will fall somewhere between "C" and "D" on your scale.

Figure 1. An example of the training materials showing film content, recorded narration, and a sample practice-booklet page. The example given is taken from the materials used for the guided group of trainees (Technique A).

the other six to treatment B. Except for the difference in the participation technique, the method of training and testing was the same for all classes. The film training was given and this was followed immediately by the test. The training and testing were presented as part of a classification battery and were administered following a rather typical test of intellectual ability. The proctors and test administrators who gave the intelligence test also ran the experimental instruction and testing session. There was no evidence that any of the trainees tested recognized that they were taking part in an experiment or perceived the experimental test procedures as anything other than a routine instruction and testing operation. During the testing, subjects were requested to report the degree of previous experience with a slide rule. All men reporting experience were excluded. The average number of years of education was approximately 12.1 years for each group. Intelligence as measured by the AFQT was nearly equal for the two groups, the average scores being 54.6 for group A and 53.5 for group B.

Measurement of Training Effects

Immediately after the showing of the film, the men were tested on their ability to read the slide-rule scales. The test consisted of 25 items. Each test item was projected onto the screen for a fixed period of time with recorded instructions for answering. The trainees' task for each item was to locate a number on a printed copy of the slide-rule scale which was similar to those used in the participation exercises. But in the test no hints or guidance were supplied to either group. In a typical test item the instruction, "LOCATE 1760," appeared on the screen and the narrator repeated, "Locate one thousand, seven hundred, sixty (one-seven-six-oh)."

About half the test items were "old" items that had been used in participation exercises or in the film itself. The other half were "new" items not previously used. By using both kinds of items it was possible to tell whether any effect of the better participation method was limited to the material actually practiced, or whether it also generalized to new problems of a similar kind. [Cf. Chapter 18.]

Results

Figure 2 shows the comparative overall results in terms of the mean number of correct answers on the 25-item post-film test. It represents a clear-cut though numerically small margin of superiority for Technique A over Technique B.

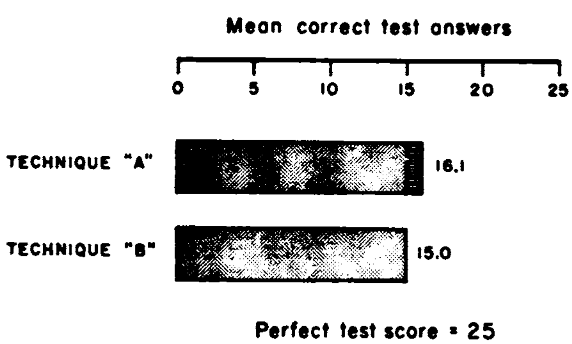


Figure 2. Mean correct test answers by technique.

The overall results show that the guidance procedure represents a positive factor leading to better learning. The relatively small size of the measured advantage does not necessarily mean that this factor is unimportant. It should be noted that in this experiment the average "baseline" level of performance for Technique B was rather high, so that the amount of room for further improvement was somewhat restricted. The mean number of correct responses on this 25-item test was 16.1 for Group A, and 15.0 for Group B. [See Supplementary Table 1.] When the reliability of the difference between these means is tested by a t test for replication means ordered by mean ability (AFQT) scores, the difference is found to be reliable.[2] ($p = 0.025$ for one tail; $t = 2.44$, 5d/f) [See Supplementary Table 2.]

[2] In this and other statistical tests used in this study, the class group (replication), composed of four quarter-flights trained and tested together, was considered the unit of sampling (with respect both to composition and conditions of testing) and, accordingly, as the unit of analysis. For the test cited here, the six experimental and six control groups were paired on the basis of mean AFQT score; hence, for the six pairs of classes only five d/f were used (each component mean being based on a class rather than on an individual or even, less conservatively, on a component quarter-flight).

An important question to be answered concerning the superiority found for Technique A is whether the guided-participation procedure improves later performance only on the problems that have actually been practiced in training, or whether it carries over to new problems as well. In separate analyses of the "old" test items (which had been used in the film instruction) and the "new" items (which had not), it was discovered that group A was superior to group B not only on the "old" items but also on the "new" items that had not been practiced. As seen in Figure 3, the measured advantages of the A over the B groups, shown by the shaded portion of the bars, appeared to be about equal for the "old" and "new" items. When the difference between the means for the new items is examined by a _t_ test (5d/f), it is also found to be reliable. (_P_: 0.01 for one tail; for the "old" items, _P_: 0.05 for one tail. [See Supplementary Table 3.])

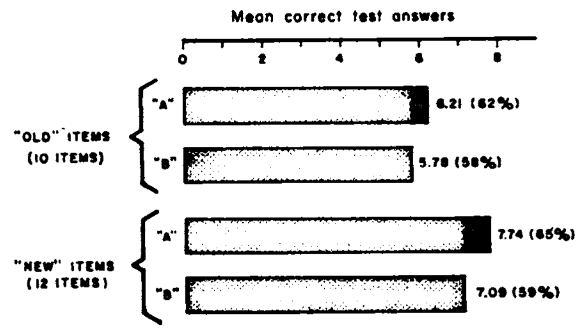


Figure 3. Mean correct test answers for "old" and "new" items by technique.

Reduction of errors resulting from guidance on the practice problems is thus seen to carry over to new problems. This is important because training is usually intended as a preparation for handling novel situations and problems, not just those encountered in practice.

An analysis of the difference in performance between the two groups was also conducted on the more difficult test items and on the easy items. The results of this analysis showed a rather large difference when the groups are compared on the 12 more difficult items but very little difference when the groups are compared on the 12 easier items. [Supplementary Table 4.] (Difficult and easy items were those below and above the median item, respectively, in overall percentage of correct answers for the two experimental groups combined.)

Thus far we have seen that Technique A is superior because subjects gave more correct test responses after being trained with Technique A than with Technique B. Put another way, this means that subjects in group A made fewer errors. We might therefore ask whether there is any systematic difference between the two groups in the nature of the error performance. To get an answer to this question, the performance of the men in each group was broken down into the following types of responses: (a) correct answers; (b) common mistakes (mistakes frequently made—for example, the number .085 is very often mistakenly located where 185 is found on the slide rule); (c) other miscellaneous mistakes.

The result of this analysis shows that the important basis for the greater frequency of correct answers in group A than in group B lies in reduction of the number of common errors rather than in the number of miscellaneous errors. Thus, 13.6 per cent of the test responses by group A fell into the "most common error" category, whereas 18.0 per cent of the responses by group B fell into this category. On the other hand, the two groups differed hardly at all with respect to all other errors (22.0 per cent for group A and 22.4 per cent for group B). The superiority of Technique A is seen to be almost entirely the result of the fact that the men in this group made fewer of the common errors than the men in group B.

Discussion

The immediate purpose of this study was to find out which of the following two audience-participation procedures is most effective:

(a) Technique A, in which participation behavior is guided so that the subject makes only correct or nearly correct responses.

(b) Technique B, in which the trainee is allowed to attempt to obtain the right answers for himself.

The evidence obtained clearly favors the first procedure (Technique A). The findings suggest the speculation that a major factor in the beneficial effect of student-participation procedures (where they are beneficial) is the fact that participation elicits the practice of correct responses, while discouraging incorrect ones. The assumption here is that practice alone is not enough. If a participation procedure led to the practice of errors, it would detract from, rather than add to, the value of a training film. It seems a good guess that some instances in which audience participation has not been found to add to the value of a film were cases in which the positive effects of practicing desirable responses are canceled by the negative effects of rehearsing incorrect ones.

In support of this view it has been seen that the benefits of guided participation can be accounted for almost entirely in terms of the elimination of common mistakes, the practicing of which this particular guidance technique was designed to prevent. It seems quite probable that in a good participation procedure incipient mistakes should be corrected before they are made.

In retrospect it seems that the question that generated this experimental study might be re-phrased. It is true that re-phrasing the question will not change the data, but perhaps an alternative way of posing the question will put the data in a more useful perspective. The original question asked, in effect, was whether guidance during participation would foster better learning. It is difficult, however, to generalize from the example of guidance that is given in this study to the construction of new guidance techniques for other dissimilar tasks. Furthermore, even if the technique could be copied for other tasks, one might ask whether or not a generalization about guided participation would hold across all possible tasks, or whether it is good only for certain kinds of learning tasks. In view of these difficulties, it might be better to put the question another way, and ask whether training is more effective under conditions that most often foster correct participation than with conditions that less often foster correct participation. In putting the question this way, it is intended that correct participation imply making correct responses to cues that will be present and reliable in the test situation. Thus, correct participation during training is not simply defined in terms of the correctness of the end response, but rather is defined jointly in terms both of the response and of cues to which the response is actually made during training. The same, or almost the same, cues must be used during training as will be used during testing if the participation is to be called correct. When the question is put this way, then we can see that for the experimental task selected, the guidance procedure that was employed was designed not only to insure that the final response during participation would be correct, but also to insure that the learners would use cues deriving from the slide-rule scale to make the correct responses, and that the cues they would use would be the same as those that must be used during testing. To see better how this is so, consider an alternative guidance technique, the use of an arrow pointing directly to the correct response. With this technique, the learner would always make correct responses, but the cues he would use might often be quite different from those he would have to use in the test situation, and this kind of guidance might not be effective.

If the data are seen as providing an answer to this alternative question, then it is somewhat easier to imagine how one might apply the findings to a different task. Thus, with a different task, it would be necessary simply to devise stimulus conditions to be employed during participation that would make it highly likely that the learner would both employ correct cues and make correct responses to those cues. [Parallels between the concepts and findings of this study and current notions of "prompting" and optimal correct-response frequency in self-instructional media are discussed in the concluding chapter. --Ed.]

SUPPLEMENTARY TABLE 1

Class Mean on Slide-Rule Test and on AFQT: by Flight (flt.), by Replication (rep.) and for Total Group. Replication Means are not Weighted for Number of Cases in Each Flight.

Replication	Flight	Group "A" Means				Group "B" Means			
		Slide-Rule Test		AFQT		Slide-Rule Test		AFQT	
		Flt.	Rep.	Flt.	Rep.	Flt.	Rep.	Flt.	Rep.
	1	17.6		63.0		13.9		52.7	
a	2	10.2	15.5	43.6	55.8	13.4	14.0	49.3	49.1
	3	16.9		51.4		14.5		50.8	
	4	17.3		64.0		14.2		43.1	
	1	14.8		42.1		15.0		48.6	
b	2	15.0	14.6	56.0	49.4	15.9	14.4	54.7	51.0
	3	13.9		50.0		14.3		52.0	
	4	14.5		51.4		12.3		48.0	
	1	18.0		62.3		20.2		67.3	
c	2	19.9	18.3	66.9	61.0	16.3	16.6	57.5	59.1
	3	18.4		55.8		16.5		60.0	
	4	16.8		56.4		13.3		52.3	
	1	14.8		59.3		17.1		57.5	
d	2	13.4	14.5	45.4	51.3	14.4	15.3	54.7	56.0
	3	16.8		42.3		13.0		52.5	
	4	13.0		56.4		16.8		60.8	
	1	15.9		51.3		15.0		52.7	
e	2	17.9	17.3	55.0	56.3	16.8	14.8	49.1	49.4
	3	18.5		67.9		15.8		50.0	
	4	16.8		50.0		11.4		46.7	
	1	17.7		53.6		14.1		60.0	
f	2	18.4	16.2	65.0	54.1	16.4	14.8	51.3	56.2
	3	14.6		37.1		15.3		64.2	
	4	13.9		60.0		13.5		51.3	
Overall Means			16.1 (64.4%)		54.6		15.0 (60.0%)		53.5

SUPPLEMENTARY TABLE 2

Statistical Test of Significance of Difference Between Means for Groups A and B. Replications are Paired on the Basis of AFQT Rank. See Table 6 for Test in Which Groups are Matched (equated) on the Basis of AFQT.

	Group A		Group B		
Rank on AFQT	Mean AFQT	Test Mean	Mean AFQT	Test Mean	Test Mean Diff.
1	49.4	14.6	49.1	14.0	0.6
2	51.3	14.5	49.4	14.8	-0.3
3	54.1	16.2	51.0	14.4	1.8
4	55.8	15.5	56.0	15.3	0.2
5	56.3	17.3	56.2	14.8	2.5
6	61.0	18.3	59.1	16.6	1.7
Mean	54.6	16.1	53.5	15.0	1.08
					t = 2.44
					p = .025 (for one tail)

SUPPLEMENTARY TABLE 3

Difference in Performance for Subjects Trained by Technique A and Technique B on Ten "Old" and Twelve "New" Items Separately. (See also Table 7.) The Data Presented are for the Replications Matched on AFQT used in Table 5.

	"NEW" ITEMS				"OLD" ITEMS		
Replication	Mean A	Mean B	Difference A Minus B	Replication	Mean A	Mean B	Difference A Minus B
1	7.28	6.70	.58	1	5.89	5.33	.56
2	7.71	7.0	.71	2	6.60	5.98	.62
3	8.58	7.19	1.39	3	6.70	5.72	.98
4	8.90	8.10	.80	4	7.45	6.90	.55
5	7.11	6.79	.32	5	5.49	5.49	.00
6	6.88	6.77	.11	6	5.12	5.28	-.16
Mean:	7.74	7.09	.652	Mean:	6.21	5.78	.425
Percentage:	65%	59%		Percentage:	62%	58%	
		t = 3.52 p = .01 (1 tail)				t = 2.45 p = .05 (1 tail)	

SUPPLEMENTARY TABLE 4

A Test of the "Significance" of the Difference Between Hard and Easy Items.*

Easier Items					More Difficult Items				
Item No.	Rank in Difficulty	(0-N) Status**	Reading	% Difference A - B	Item No.	Rank in Difficulty	(0-N) Status**	Reading	% Difference A - B
1	25	(N)	147	-1.9	2	2	(N)	0.74	12.6
4	16	(N)	1420	.8	3	6	(N)	26	13.1
5	19	(N)	445	1.8	6	8	(N)	0.394	10.5
7	21	(N)	0.114	-1.5	8	12	(N)	314	10.5
12	15	(N)	245	9.4	9	10	(N)	18	5.3
13	18	(N)	109	3.3	10	1	(N)	0.042	8.9
16	22	(*)	1.32	-2.3	11	4	(*)	2.3	4.5
20	14	(0)	212	7.3	14	5	(0)	0.085	13.5
21	24	(0)	415	3.4	15	9	(0)	313	5.9
23	20	(0)	1760	- .6	17	7	(0)	23	5.0
24	23	(0)	0.151	-1.4	18	3	(0)	0.82	9.0
25	17	(*)	1.02	-3.7	22	11	(0)	0.272	7.5
Mean				1.22	Mean				8.85

*The value of t for hard versus easy calculated from these data is 5.13 (P = .01, two tails). Performance of total groups on the various test items are the basic entries here, with degrees of freedom based on number of items; that is, the "significance" evaluation is, strictly speaking, a check on errors due to sampling of items.

**The "0" or "N" means "old" or "new" item; items marked with an asterisk differed only in placement of decimal point as between instruction and test.

CHAPTER 16

THE EFFECTIVENESS OF INSTRUCTION IN READING A SCALE AS INFLUENCED BY THE RELATIVE AMOUNTS OF DEMONSTRATION AND PROBLEM-SOLVING PRACTICE[1]

Gregory A. Kimble
J. Jepson Wulff

It is generally agreed that for learning to take place, responses of some kind must be made by the learner (Cf. Chapter 1). It follows from this that the use of training procedures that encourage responding by the learner should increase the amount learned from training. To test this inference, a number of studies have been done in which the teaching effectiveness of audience-participation techniques was investigated (e.g., Hovland, Lumsdaine, and Sheffield, 1949; Michael and Maccoby, 1953; May and Lumsdaine, 1958). In general these studies have shown that the amount learned is increased by the use of participation techniques.

In a training situation, however, not all learning is relevant; only certain responses of all those which might be learned are called "correct." With this restriction, just any participation on the part of the learner is probably not a sufficient condition to establish correct performance. Practice of correct responses must be encouraged (Cf. Chapter 15). Yet further, it may be necessary to ensure that during participation the correct responses are made to appropriate stimuli. Thus it may be advantageous to utilize participation techniques that not only encourage the learner to make responses, but also make it highly likely that he will make the correct responses to stimuli that will be useful stimuli when the learner subsequently is placed in a practical performance situation. This emphasizes that training conditions should ensure: (1) that the learner will use performance-relevant stimulus features; (2) that his responses will be those called "correct" for the task; and (3) that he will make these correct responses to the appropriate stimuli.

One way to approximate a training device that needs these requirements is in common practical use. It is done by pointing out to the learner the particular stimulus features that are thought to be useful cues for the desired responses, and by

[1] The experiments reported in this chapter were performed in 1952 as part of the work under Contract AF 18(600)-35 between the Air Force and Yale University. The authors were assisted in this work by Doris G. Kraeling and William McGuire; see also Footnote 1 to Chapter 15. This report is a revision of an unpublished Staff Research Memorandum prepared by Dr. Wulff for the AFPTRC Training Aids Research Laboratory (Kimble and Wulff, 1954), based on an earlier contract report written at Yale by Dr. Kimble.

demonstrating the responses that will be called correct.[2] This method assumes that the learner will participate, covertly at least, without specific instruction to do so. The method has two potential shortcomings: (1) the learner may err by coming to depend upon the purely artificial devices used to point out the relevant stimulus features as pertinent cues in themselves; and (2) he may not be provided with the optimal conditions for practicing the correct responses to the stimulus features.

Training procedures used to avoid these shortcomings require overt participation by the learner during training. The learner does not, however, always maximize the probability that the correct responses will be made to useful cues during training. If the participation exercise simulates actual performance conditions (i.e., provides no aids not present in the performance situation), the learner may be left to his own devices in selecting the stimulus feature to which he will connect his response. Unless he has been adequately prepared, the participation exercise may provide an occasion for him to practice the correct responses to irrelevant cues, making the chances of subsequent success poorer. If, on the other hand, the participation exercise provides too much guidance, it may encourage practice in how to use the guides, rather than practice in making correct responses to useful stimulus features.

Conjectures of this kind about the learning that might occur during audience participation led Kimble and Wulff (1953a,b) to undertake a study in which a modified participation technique was used (see Chapter 15). The aim was to make it highly likely that learners would utilize useful stimulus features during participation. This technique was labeled "response guidance," although from the standpoint of the present analysis "guidance in stimulus selection" might be more appropriate. The training task was the same as the one used in the present experiment (learning to read the "C" and "D" scales of the slide rule). First, common types of errors were determined by preliminary studies, and this information was used to make up a number of "guided participation" problems to be used in training. These problems required the learner to locate given numbers on a printed slide-rule scale, and guidance was provided by placing "guidance" lines on the slide-rule scale (see Fig. 1 of Chapter 15), in such a way as to exclude common errors and yet to include as wide a range on the scale as possible. An alternative form of guidance was provided in some cases by showing on the printed scale the correct location of a different problem of the same type as the one the learner was to solve. The fact that the response-guidance procedure was found to be significantly superior to unguided participation as a training method might lead to the inference that the guidance technique was successful in getting the learner to use stimulus features that proved useful during later tests. Alternatively, it could be inferred that the effect was due merely to the fact that the "guidance" group made more correct responses during participation (i.e., that the superiority reflected only a practice effect).

[2] The training problem used in the present experiment—learning to read the slide-rule scale—is one for which most adult learners will have in their repertoire all the prerequisite responses necessary for adequate performance. Thus in this task, as in most other cognitive-learning tasks, probably little or no learning of overt responses is required. The present discussion is intended to cover only such tasks, and a consideration of response learning would unduly complicate the presentation.

It seems likely, however, that use of the correct stimulus must come before gains can be realized from practice. The present experiments were designed to test the hypothesis that the training effectiveness of participation techniques is due in part to conditions that force the learner to use stimulus features that will be useful in a performance situation.

The design of the present experiment utilizes several alternative training programs, all of which teach the reading of the slide-rule scales, and all of which contain the same number of examples, but which differ in the proportion of the total training devoted to guided participation. Thus, for example, if a program is to contain ten practice examples, all of these may be presented with the answers given, or half of the examples may be presented as participation exercises, or again, all ten of the examples may be participation exercises.

The actual proportions of the practice examples devoted to participation in the first experiment which follows are 0/16, 8/16, and 16/16. Four additional examples were used in a basic-training part of each training program. These four examples were always presented as demonstration examples. In the second experiment reported, the proportions were 0/16, 4/16, 8/16, 12/16, and 16/16.

With this task[3] if the most benefit were to accrue merely from practicing the correct response (whether overtly or covertly), the no-participation condition should be superior, since the correct response was always shown with instructions to study it. If motivation is proportional to the amount of participation, or if the effectiveness of participation is due merely to the fact that it provides practice in how to perform, the highest proportion of participation should be superior. If, however, it is most beneficial to provide the minimum of training with complete guidance necessary to point out the relevant stimulus features, and then to require that these stimulus features actually be responded to in a situation that forces their use, then an intermediate proportion of participation may be superior.[4]

---

[3]As noted above, this discussion is relevant primarily to tasks involving little or no response learning. It should further be pointed out, however, that it is not intended to cover all non-motor tasks, since, for example, the problems involved in constructing a training device would be quite different for a task involving stimulus features that were completely familiar to the learner. The present analysis is intended to refer to tasks in which the useful stimulus features that must be discriminated are obscure or are highly similar within the task.

[4]This is not intended to imply that the 50 per cent participation will always be superior. It implies only that the optimal proportion of participation will usually be less than 100 per cent and greater than zero per cent. (A similar hypothesis is also considered in a somewhat different context in the paper by Angell and Lumsdaine, Chapter 24.) The precise proportion will depend upon the total number of both kinds of examples used and the difficulty of the task.

Experiment I

Method

Experimental Materials. Trainees in this experiment were taught how to read the "C" and "D" scales of the slide rule by means of a twenty-minute recorded narration used in conjunction with individual booklets of printed examples that were cued to the recording. The instruction may be broken down into two parts, a basic part which was the same for all trainees, and a part which differed for each experimental group. The basic part consisted of the entire recorded narration, written in a way that was appropriate for all experimental groups, and of four examples presented as demonstration examples in all of the booklets. Thus the narration was the same for all trainees; differences among the treatments given the three experimental groups were introduced only with sixteen practice examples spaced throughout the printed booklets. For these examples the narration instructed the trainees to turn to the appropriate page, and instructions about how to do the problem were printed on the page.

For each of these problems the instructions were either (a) to study the solution to the location of a given number shown on the printed scale, or (b) to draw a vertical line to show the location of the number given. If the latter instruction were given, the response guidance method (cf. Fig. 1, Chapter 15) was always used. Three different forms of the booklet were prepared, each with different proportions of participation and demonstration examples among the sixteen. Each one of the following proportions of participation examples determined an experimental group (or training condition): 0/16, 8/16, and 16/16. The 0/16 group may be called a "no-participation" group. The 16/16 group, however, was not strictly a "100 per cent participation" group, inasmuch as trainees under this condition received the same four basic no-participation examples as all other groups.

Design of the Experiment. Because the differences in treatment appeared only in the booklets, it was feasible to use what was substantially a randomized block design in the administration of the experiment. Air Force basic trainees served as subjects in classes of about 45 men each. Each class was seated alphabetically in five "columns" of seats. The men marched into the room in alphabetical order and filled one column of seats at a time. The three different versions of the booklet were distributed among these seats so that the first three rows of men (one block) received booklets for one condition of training (N=15); the second three rows received booklets for another condition of training, and the third three rows received booklets for the third training condition. The assignment of experimental materials to blocks of trainees was random. Six classes of men were used, comprising six replications; each class was trained at a different time.

Procedure. After a general introduction to the task, the trainees were instructed to listen to the recording and to follow the narration by looking at the appropriate pages in their booklets, making sure to turn the pages as they were directed to do so by the recording. Whenever one of the sixteen practice examples was referred to, the narrative stated, "Turn to page X and follow the directions printed on that page." This instruction applied equally well whether a trainee's example was a participation problem or a demonstration example. A period of 30 seconds was allowed for the subject to draw a line to locate the number given on the page (if the example was of the participation type) or to study the example (if of the demonstration type). In either case, the number used in the example was the same.

This training procedure was followed immediately by a 25-item multiple-choice test. Each item was answered by selecting that one of five lines drawn on a printed slide-rule scale which indicated the correct location of a given number. The test was paced by means of a recording with 30 seconds allowed for each problem. All the numbers used in the test were different from the numbers used as examples in training.

Results

The principal results are presented in Table 1 and are summarized in Figure 1. This figure shows the mean scores for each group and the "one per cent confidence limits" about these means.[5] It can be seen from the figure that the 8/16 participation group achieved a higher mean test score than the 16/16 participation group. The difference of 2.25 between these two means is reliable at the .01 level of confidence for one tail. It can also be seen from the figure that, although the mean for the 8/16 group is higher than that for the no-participation group, the difference between these means does not approach significance.

The size of each block in the experiment (N of about 15 for each block) was not enough to achieve stability of the mean of the technical-specialty (ability) scores for each block (see Table 1). A second analysis was therefore performed on samples selected from each treatment group without respect to replication, such that the distribution of ability scores was the same for each group. (See Table 2.) The difference between the mean for the no-participation group and the mean for the 8/16 group is larger when the groups are equated in this manner. The difference of 1.22 is not reliable, however ($t<1.0$). The difference between the 8/16 group and the 16/16 group (2.69) is still reliable (at the .03 level of confidence for one tail) in this analysis.

[If the data for 0/16 and 16/16 were pooled, the difference of about 1.9 between the joint mean for these groups and the 8/16-group mean would approach significance, in support of the general hypothesis that "an intermediate level of participation" should be superior to the extremes represented by 0/16 and 16/16. The breakdown

[5] [These "one per cent confidence limits" are based on the S. E. of the overall mean for each treatment (about .67, as shown in Table 1); using this as a normal deviate gives, for p of .005 (each tail), bars extending 1.75 score units from the mean. The confidence limits thus apply to a mean based on random samples of about N=88 individuals (see Table 1), rather than specifying the expected range for replication means in the present design.]

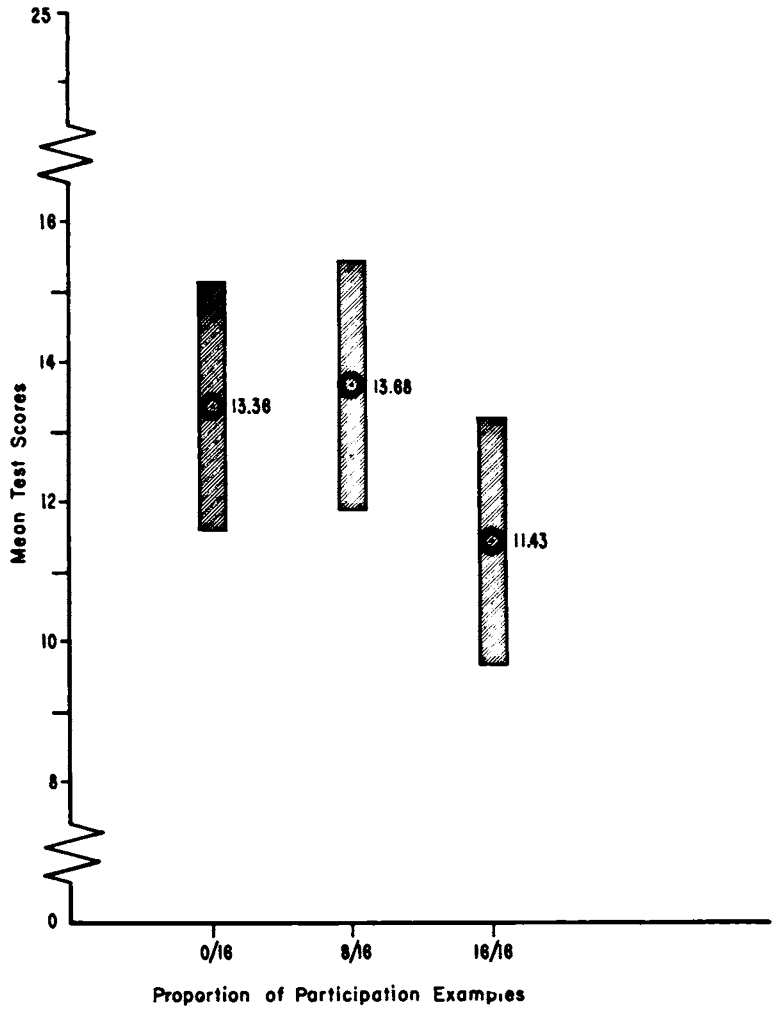


Figure 1. Mean test score for each treatment group in Experiment I. The shaded bars about each mean indicate the range within which sample means would be expected to fall 98 per cent of the time if the true (population) mean were the same as the mean obtained in this experiment (see Footnote 5).

by ability presented in Table 2 also shows one further revealing fact. This is that the relative advantage of the 50 per cent participation condition does not seem to be uniform for different levels of ability. For example, if we group together some of the ability categories so as to obtain four approximately equal-sized subgroups—"low" (categories 1, 2, and 3), "medium low" (categories 4 and 5), "medium high" (categories 6 and 7), and "high" (categories 8 and 9)—it becomes apparent that the largest differences among treatments, at least with respect to the superiority of the 50 per cent over 100 per cent treatments, is found in the "medium" groups, which together show an average score of about 12 for the 0/0 treatment and around

TABLE 1

Mean Test Scores (and Technical-Specialty Ratings) for Each Treatment Group by Replication Based on all Subjects (Experiment I)

	Proportion of Participation Problem for each Group								
	0/16			8/16			16/16		
Replications	Mean Test Score	N	Mean Tech Sp. Rating	Mean Test Score	N	Mean Tech Sp. Rating	Mean Test Score	N	Mean Tech Sp. Rating
1	14.67	15	6.93	14.47	15	6.50	15.53	15	5.20
2	13.14	14	4.46	14.73	15	6.13	12.00	14	5.79
3	12.73	15	5.40	14.46	13	6.23	11.53	15	5.93
4	13.47	15	6.33	12.87	15	5.93	9.70	13	4.92
5	14.64	14	5.79	12.27	15	4.93	8.40	15	5.00
6	11.60	15	4.80	13.40	15	4.71	11.20	15	4.86
	0/16			8/16			16/16		
Overall Mean Test Score $\frac{(\Sigma X)}{n}$	13.36			13.68			11.43		
Total N	88			88			87		
σ Mean	0.674			0.674			0.678		
Difference between Means			0.32			2.25			

$\sigma M_1 - M_2 = 0.956$; $t = 2.35$; $p = .01(.02)$

TABLE 2

Mean Test Scores for Each Treatment Group (Experiment I) by Technical-Specialty Category Matched with Equal N's for Each Treatment Group

Technical Specialty-Rating Category	N (per treatment)	Proportion of Participation Problems for each Group: 0/16 Mean Test Score	8/16 Mean Test Score	16/16 Mean Test Score	Overall Mean for T-S-Rating Category
1 and 2	6	4.67	8.67	7.33	6.89
3	6	9.33	8.67	6.33	8.11
4	8	10.25	12.12	5.38	9.25
5	9	10.67	12.44	6.89	10.00
6	10	13.30	10.00	8.90	10.73
7	8	14.75	18.62	14.75	16.04
8 and 9	15	17.87	19.67	19.73	19.09
Overall Mean Test Score		12.60	13.82	11.13	
σMean		1.01	1.01	1.01	
Difference between Means		1.22 (0/16 vs. 8/16)	2.69 (8/16 vs. 16/16)		

$\sigma M_1 - M_2 = 1.42$; $t = 1.90$; $p = .03 (.06)$

13 for the 8/16 (50 per cent) treatment, but of only about nine for the 16/16 treatment. By contrast, the "high" group shows relatively little difference among any of the treatments, and though the "low" group appears to peak in the middle as compared with either end, the differences among treatments for this sub-group are relatively slight—being at most only about two points (as compared with an average difference of about 4-1/2 points between the 8/16 and 16/16 treatments for the "medium" groups). These facts might be interpreted to reflect the notion that a trainee of very low ability does not learn much from a training device, no matter how it is presented (despite a tendency for a middling amount of participation to be superior to either extreme of zero or maximum), while a very able man, on the other hand, learns a great deal from the presentation, however it is implemented. Following this line of thought, we might further suggest that it is in the middle range of ability that treatment differences are likely to show up most conspicuously, as particularly evidenced here by the fairly substantial margin of superiority for the 8/16 over the 16/16 for the medium-ability groups.--Ed.]

Experiment II

Method

In all respects except the following, the materials, design, test, and administration of this experiment were the same as for Experiment I.

Five conditions of training were compared in Experiment II. The proportions of "participation examples" used to make up the five conditions were: no participation, 4/16 participation, 8/16 participation, 12/16 participation, and 16/16 participation. The booklets for the no-participation, the 8/16, and the 16/16 participation groups were the same as those used in Experiment I. The 4/16 and the 12/16 booklets were prepared according to the same plan as that used for the other booklets.

Four replications were run. For each replication the treatment blocks were assigned positions as determined by a table of random numbers except for the no-participation condition. The no-participation booklets were always placed at the rear of the room in this experiment.[6]

Results

The principal results of Experiment II are presented in Table 3 and are summarized in Figure 2. This figure shows the mean test scores for each treatment group and the "five per cent confidence limits" about each mean in which the true mean would be expected to fall (cf. Footnote 5, supra). It can be seen from the figure that all of these ranges overlap considerably. The difference between the no-participation mean and the mean for the 4/16 group of 1.72 is not reliable ($t = 1.30$)—see Table 3. It seems apparent from this fact and from the figure that none of the other differences approach reliability.

Inspection of the participation booklets revealed that about one-third of the subjects had completely failed to follow instructions to participate. Therefore, a second analysis was performed with means based on test scores for subjects who actually participated on at least several of the participation problems. Replications were ignored in this comparison. This selective procedure increased the differences among the treatment means. However, the largest difference, the difference between the mean for the 12/16 group and that for the 8/16 group (2.16) is not reliable ($t = 1.33$)—see Table 4. None of the other pairs of adjacent means are reliably different.

[6] This was done so that "short" flights would always lessen the N of the same training group. It was felt at the time that the no-participation condition was of least interest, and thus this condition was chosen to occupy the block of seats at the rear of the room to facilitate shortening columns so that all other blocks would receive a full complement of 15 men each.

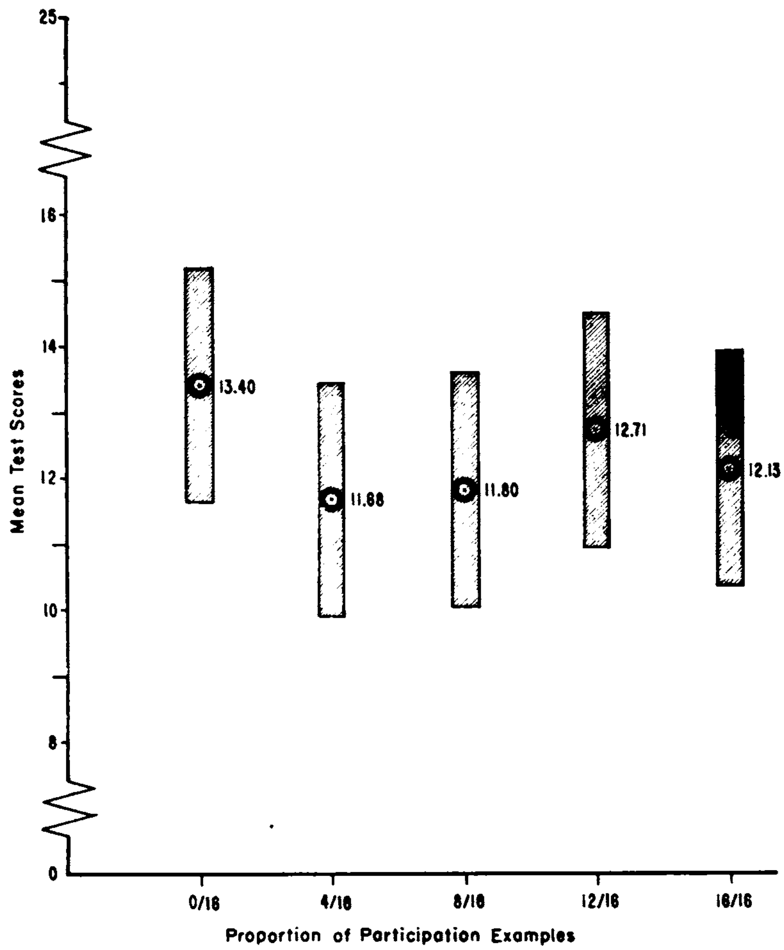


Figure 2. Mean test score for each treatment group in Experiment II. The shaded bars about each mean indicate the range within which the sample mean would be expected to fall 90 per cent of the time with a true mean equal to the one here obtained (cf. Footnote 5).

Discussion

The results of these experiments are in tenuous agreement with the expectation that an intermediate proportion of participation exercises is superior in teaching effectiveness to a very large or very small proportion of participation exercises. In Experiment I it was found that 8/16 participation was reliably superior to 16/16 participation. In Experiment II the pattern of the treatment means conformed to the expected pattern, with a high point for 12/16 participation when the non-participators were excluded from consideration, but the differences between the means were not reliable (p about .10, one tail). In neither experiment, however, was the no-participation condition reliably inferior to the intermediate mode.

TABLE 3

Mean Test Scores for Each Treatment Group by Replication
Based on All Subjects (Experiment II)

	Proportion of Participation Examples for each Group														
	0/16			4/16			8/16			12/16			16/16		
Replication	Test Mean	N	Educ* Mean	Test Mean	N	Educ Mean	Test Mean	N	Educ Mean	Test Mean	N	Educ Mean	Test Mean	N	Educ Mean
1	16.83	6	11.83	10.86	15	10.80	11.46	15	11.27	14.40	15	10.20	14.00	14	11.86
2	10.27	11	10.18	13.06	15	11.20	12.57	14	10.43	13.57	14	10.60	11.66	15	10.27
3	12.72	11	11.55	13.00	15	10.33	13.93	15	11.33	13.06	15	11.00	12.93	15	11.07
4	15.16	12	10.67	9.80	15	10.20	9.26	15	9.93	9.86	15	10.60	10.06	15	10.93

	0/16	4/16	8/16	12/16	16/16
Overall Test Mean $\frac{(\Sigma X)}{N}$	13.40	11.68	11.80	12.71	12.13
N	40	60	59	59	59
σ Mean	1.03	.837	.844	.844	.844
Difference between Means		1.72	0.12	0.91	0.58

*Education Mean is mean of total number of years of schooling for each subject.

TABLE 4

Mean Test Scores for Each Treatment Group in Experiment II
Based Only on Subjects Who Participated as Instructed

	Proportion of Participation Problems for each Group						
	4/16		8/16		12/16		16/16
Mean Test Score	11.54		11.29		13.45		12.59
N	33		24		31		29
σ Mean	1.04		1.23		1.07		1.11
Difference between adjacent means		0.25		2.16		0.86	

Note: No comparable selection of participators could be made for the 0/16 group, which is therefore not included in this table.

The relatively high means for the no-participation condition may have been due to poor sampling with respect to general ability, or, in the case of Experiment II, to the small number of subjects within blocks. On the other hand, the pattern of means for the second experiment (shown in Figure 2) may imply that a small amount of participation detracts from the effectiveness of the teaching device, and that it is only with a relatively large amount of participation that the positive effects of participation counterbalance the negative effect. The relatively large variance of the treatment means in those experiments makes it impossible to do more than conjecture about this.

Three major uncontrolled factors operated in these experiments. It was not possible to control adequately the extent to which subjects followed instructions to participate. It was evident from the individual participation booklets for Experiment II that many subjects did not participate when instructed to do so. It is also apparent from the booklets that many of the no-participation subjects did participate actively without instruction to do so. Since the hypothesis under examination was not concerned with the practical utility of the specific instructions used in the experiment, these failures to follow instructions make it difficult to determine treatment test means that accurately reflect the true means for the treatments investigated.

In addition, it was difficult to identify subjects with prior slide-rule experience, and for that reason subjects were included in the analyses without respect to consideration of their prior experience. All subjects were instructed to indicate the extent of their prior experience on the test booklets, but in many cases there were indications that these responses were not valid.

Finally, the distribution of subjects with respect to ability was not stable from block to block (see Tables 1 and 3). It is known from prior experimentation (Kimble & Wulff, 1953a, b) that test scores for the slide-rule task are correlated with ability (r = approximately 0.60).

Since the conformity of the treatment means with the expected pattern of these means tended to improve when controls for two of these factors were imposed on the data, one might conclude that the statistical unreliability of the present findings does not reflect a true "no difference." Rather the data could be held to justify the expectation that improved experimental methods might better demonstrate the expected relationship between proportion of participation exercises and teaching effectiveness.

CHAPTER 17

THE EFFECT OF ANIMATION CUES AND REPETITION OF EXAMPLES ON LEARNING FROM AN INSTRUCTIONAL FILM[1]

A. A. Lumsdaine
Richard L. Sulzer
Felix F. Kopstein

Introduction

This chapter consists of two main sections, both based on the data from a single multi-factor experiment. For the sake of clarity, the methods and findings appropriate to two major forms of experimental variation have been separately described in the two sections. Section I deals with the effect of some animation techniques that were manipulated in the experimental films, and Section II deals with

[1]The study reported in this chapter was conducted as an in-service project at the U. S. Air Force's Human Resources Research Laboratories in 1950-51. Previous reports by Lumsdaine and Sulzer (1951) on the findings reported in Section I of the present chapter, and by Sulzer, Lumsdaine, and Kopstein (1952) on the findings reported in Section II, were issued as publications of the Human Resources Research Laboratories but are of very limited availability. The present chapter was prepared by the editor on the basis of these earlier reports, for the writing of which he and F. F. Kopstein were primarily responsible. Results of this study and a related study were also prepared as a short "film report" on 16 mm. sound film, which was initially presented at the 1950 convention of the American Psychological Association.

Except for a few of the more detailed supporting tabulations, all of the essential information contained in these reports is presented in the present chapter. The experimental tests on which this report is based were conducted by R. L. Sulzer and M. Jacobs at Lackland Air Force Base, through cooperation with the Directorate of Personnel Research, Human Resources Research Center, in June, 1950.

Appreciation is expressed to George Owen and other members of the HRRL film-production staff for their assistance in preparing the experimental film materials, and to Karl D. Kryter, then director of HRRL, for general administrative support for the program under which this study was conducted. R. L. Sulzer is now on the staff of the Operational Applications Laboratory of the Air Force Command and Control Division, Air Research and Development Command. F. F. Kopstein, until recently on the staff of the Training Research Branch of the Aero-Space Medical Laboratory, ARDC, is now with the Burroughs Corporation.

the effects of repetition of instructional examples as manipulated experimentally both in the films themselves and in supplementary instructional materials that were used in conjunction with them. Several supplementary tables of data applicable to one or both of these sections are appended at the end of the chapter.

Section I. The Effect of Animation Techniques

The study reported in this section deals with one specific aspect of the general question of what can be added to the training value of film material through the use of motion picture animation techniques. "Animation" techniques are, of course, many and varied. This report does not deal with the more complex sorts of animation, such as the figure-animation found in the animated cartoon. Rather, it is restricted to the study of simple "pop-in" labels superimposed over the pictures, moving arrows to direct attention of the student to successive salient parts of the pictorial material, and the like.

Method

Instructional Materials

Animated Films. The kind of animation effects studied can be characterized in terms of an illustrative excerpt taken from one of the experimental films used in the study. This portion of the film used an example to explain how to read a micrometer setting. It employed moving arrows, pop-in labels, etc., to match the narration. The basic illustration at the beginning of this excerpt is an extreme close-up of the micrometer scale, over which the accompanying narration begins as follows:

> "[The reading] is determined by first referring to the scale on the barrel. The figure seven" ...(arrow pops in to point to the "7" on the barrel)... "indicates seven hundred thousandths." (Arrow moves over to one side of the micrometer and ".700" pops into the place where it now points.) "Two more graduations are seen." (Two arrows move in, to point to these.) "That's another .050, which is fifty thousandths..." (The two arrows move over, fuse and point to ".050" as it pops in under the ".700.")

The experimental film used in the study was adapted from a standard training film on the micrometer, produced during World War II by the U. S. Office of Education, and supplemented by additional footage shot specifically for purposes of the study. The experimental film also contained considerable material on the construction and care of the micrometer, but the present study focused mainly on the comparative effectiveness of different versions of the film in teaching men how to read micrometer settings. This instruction was conveyed primarily via the presentation of several examples of the kind illustrated above. (The edited films used for experimental purposes covered only the use of the outside micrometer, thus omitting some of the material contained in the original USOE film from which they were adapted.)

All films contained a short introduction explaining the purpose of the micrometer, its nomenclature, and the basic principles of its construction and use. This introduction was followed in each film by three or more examples of micrometer reading, showing how to read various settings of the instrument. These examples were inserted between interstitial or "enrichment" sections telling about the technique of using the "mike," how to care for it, malpractices to avoid in its use, how to check its accuracy, and so forth. The variation in "animation" versus "non-animation" was employed primarily in the examples that showed how to read the micrometer. In certain other portions of the presentation—for example, in explaining the principles of the micrometer's operation—identical animation techniques were used in both the animated-example films and the non-animated-example films.

Comparative Non-Animated Films. The animated examples used in one set of experimental films were exactly paralleled in non-animated films. The instruction of micrometer reading in the non-animated films was identical with that in the animated films except for the absence of all animation (pointing, labeling, etc.) in the instructional examples. The corresponding animated and non-animated films used the same sound track and, except for the variation in presence or absence of animation effects, were substantially identical, frame for frame. Given these two sets of films, alike except for specific animation effects, the net contribution of the animation could be isolated and assessed. The experimental test procedures used for this purpose are outlined briefly below.

Experimental Design and Procedure

Design. The comparison between the films using animated examples and those using non-animated examples was actually made under eight different experimental conditions, with two class groups of students (Air Force basic trainees) using animated films and two using non-animated films for each of eight conditions. The eight conditions represented all eight possible combinations of these three factors: (a) whether or not a pre-film test on micrometer reading was given before the film; (b) whether the film presented six examples of micrometer reading, or only three examples; and (c) whether the film was or was not followed by a supplementary sound slide-film giving additional examples and pointers on how to avoid the commonest types of error in reading the micrometer. (Results of analyses to determine the effects of the latter two factors are presented in Section II.) The combination of these three factors, plus the factor of animation versus non-animation, resulted in a 2x2x2x2 factorial design, requiring 16 class groups for a single replication. (Administrative considerations dictated the use of pre-formed class groups, rather than individuals or groups composed by random sampling, as the units for experimental instruction, testing, and hence analysis.) Two complete replications were tested, with all cells of the first replication being completed before the second replication was started. In the first replication each class group was a half-flight (about 25 to 30 men); in the second replication, full-sized flights (about 50 to 60 men) were used. Within each replication the assignment of animated versus non-animated films to successive class groups tested was determined by chance through the use of random-number tables.

Measurement of the Instructional Outcomes of the Films. Tests used to assess the students' ability to read micrometer settings consisted of a series of slides, each showing a micrometer setting. Each slide was projected on a screen for a fixed interval of time, and the men were instructed to write down the value of the setting on a test blank. (Previous experimentation had indicated that if men could read pictured settings correctly they could also read the settings when given an actual micrometer to read.)

Five different settings were used to test the students' prior (pre-film) ability in reading the micrometer; fifteen different settings were used to test micrometer-reading ability after exposure to any one of the experimental films (see below). In addition, other test questions were used to test effects of the film on knowledge of the nomenclature, construction, and principles of the micrometer.

Experimental Training Procedure. Each group of men was shown one of the films and was tested immediately afterward. Film showing and testing were done in typical classroom buildings at an Air Force basic-training (indoctrination) center. Procedures and conditions of showing for both animated and non-animated films embraced considerable range and variety. In part this variability was deliberately introduced, for the purpose of increasing the generality of application of the results —that is, to avoid the possibility that the findings might apply only to particular, limited circumstances. But all controlled variations in conditions and procedures for the animated films were strictly paralleled by corresponding variations for the non-animated films. This practice, as well as other more technical aspects of the experimental design and procedure, was designed to assure an unbiased comparison of the effects of the two sets of films.

In all cases participants were told prior to the film showing that they would be tested after seeing the film. Carefully standardized instructions were used in all phases of instruction and testing; these instructions were read by specially trained proctors. Testing followed immediately after the conclusion of the instruction. In the post-film tests, as well as in the short pre-film tests given to half of the classes, each test-slide (micrometer setting) was shown for 30 seconds. The settings used in the pre-film test were .040, .396, .725, .950, and .603. The settings used in the post-film test are shown in the first column of Supplementary Table 1, at the end of the chapter. On the tests, men were required only to write the correct total reading for each setting; they did not have to write down the component parts of the reading. Knowledge of nomenclature was tested by presenting a single slide on which five key parts of the micrometer were indicated by numbered arrows, and directing men to write in the names of these parts on a test blank, on lines numbered to correspond to the five parts indicated in the slide.

Processing and Analysis of Data. Test data were punched into IBM cards and were machine-tabulated for accurate and efficient analysis. Results were double-checked to ensure accuracy. In addition to specific comparisons between experimental sub-groups, a complete analysis of variance was performed on the basis of the class means for micrometer-reading scores for the 32 class groups tested, replication variance based on the observed variability of 16 pairs of means being used as a measure of error. In all analyses, either for the populations as a whole or for sub-populations based on intelligence, etc., class means were used as basic observations, with degrees of freedom based on the number of separate classes (showings) rather than on the number of individuals. This meant that the analysis was much more rigorous than one using N as the number of individuals, since any selective sampling factors, and variations in showing conditions from class to class and/or day to day, were taken fully into account in the error estimate. Further checks were also imposed by applying non-parametric tests (Wilcoxon's summed-ranks method) in addition to the standard normal parametric statistics for small samples.

Results

Initial Ability

Prior to seeing the film, very few of the men could read the micrometer at all. On the pre-film test, 93 per cent of the men failed to answer any of the five test-items correctly. Training thus started virtually from scratch. The average pre-film test performance was only three per cent correct answers. Pre-film performance was comparable for the animation and non-animation groups. (Average scores were 2.6 per cent correct and 3.6 per cent correct, respectively.)

Gains Produced by the Films

Because of the very low level of pre-film knowledge or skill, test performance following the film is almost entirely attributable to gains produced by the film. On the average, half of the men got scores of 60 per cent or better on the micrometer-reading test after seeing one of the films. The more intelligent men, as will be seen later, did considerably better than this. The average percentage (for all films) of fully correct answers given before and after the film showing was three per cent and 50 per cent, respectively, so that the average gain from the film is represented by a mean score increase of 47 per cent correct answers. A significant difference was found between the pre-film-test groups and the groups which did not receive the pre-film test. The mean for the former was 8.08, and for the latter, 6.86. (See Supplementary Tables 2, 3, and 4.)

Advantage Achieved by Animation Devices

The comparative levels of micrometer-reading skill achieved with the animated and non-animated film versions were as follows, as expressed in terms of the average percentage of correct answers on the micrometer-reading test:

Animated films: 54.6 per cent correct
Non-animated films: 45.0 per cent correct

The added gain produced by animation was shown by the highly significant difference of 9.6 per cent, which represents about 17 per cent of the possible room for improvement above the level attained by the non-animation film. The 32 overall class means for scores in the micrometer-reading test given after the four major experimental treatments (results based on test data for 1,283 men) are presented in Supplementary Table 2, which also shows the corresponding class means for high AQE (Airmen Qualifying Exam) and lower AQE sub-groups (data for the 1,267 men for whom AQE data were available).

These data give one of several possible ways of representing the comparative value of the animated and non-animated films. The clear superiority of the animated films is also seen in comparing the two sets of films in terms of the percentages of men who achieved various criteria of measured proficiency, as summarized in Table 1. For example, 59 per cent of the "animated" group got more than half of the test questions correct, while only 46 per cent of the "non-animated" group did this well.

TABLE 1

Cumulative Percentages of Individuals for Animated and Non-Animated Films, and Cumulative Number of Animation and Non-Animation Class Groups, Attaining Mean Scores of Indicated Size on Micrometer-Reading Test

No. of test items correct	(a) Cumulative per cent of individuals attaining or exceeding each score level			(b) Cumulative number of class groups with overall mean scores at or above each level		
	Anim.	Non-Anim.	(Diff.)	Anim.	Non-Anim.	(Diff.)
14 or more	21	22	(-1)	--	--	(--)
12 or more	36	32	(4)	--	--	(--)
10 or more	51	41	(10)	4	1	(3)
8 or more	59	46	(13)	9	6	(3)
6 or more	66	50	(16)	13	11	(2)
4 or more	70	54	(16)	16	14	(2)
2 or more	78	66	(12)	16	16	(0)

Two things are evident from the results. First, the numerical measure that may be cited to indicate the margin of superiority for the animated films varies somewhat, depending on how the results are expressed. Thus the absolute values of the percentages themselves should not be taken too seriously. Second, any way the pie is sliced, a clear and substantial margin of superiority is shown to result from the use of the animated films.

Statistical Significance and Consistency of Results

Analysis of variance and other statistical tests (see Supplementary Table 3-A) show that the superiority of the animated films is highly reliable—at better than the 99 per cent level of confidence ($P < .01$). The reliability of the results can also be gauged from the fact that a number of different showing conditions were used in the study, and that the films using animated examples were superior in every one of eight experimental conditions employed. Specifically, the animated films produced better learning than the non-animated films: (1) regardless of whether or not the film was preceded by a preliminary quiz orienting participants to the nature of the task to be learned; (2) regardless of whether longer films (with six instructional examples) or shorter films (three instructional examples) were used; (3) regardless of whether or not a follow-up supplementary slide-film was used following the film showing; and (4) under all eight possible combinations of the above three factors. (The complete data from which these assertions are derived are presented in Supplementary Table 2, at the end of the chapter.) The animated films were also superior to the non-animated films regardless of: (5) whether the films were shown under relatively favorable temperature conditions (morning showings) or under extremely hot, humid conditions; (6) variations in the size of the audience and distances men sat from the screen; and (7) whether proficiency in micrometer-reading after the film was strictly scored (fully correct answers) or more leniently scored (credit given for partially correct answers). Data underlying these latter assertions are presented in Supplementary Tables 1 and 2 at the end of the chapter.

Audience Characteristics as Factors Influencing Results

Results were separately analyzed for men of different educational background and differing mental ability as measured by the Airman Qualifying Examination (AQE). The superiority of the learning produced by the animated-example films over a considerable range of mental ability and educational background is evident from Table 2. The brighter and better educated men naturally learned more than average from both the animated and non-animated films. (For example, the average scores for the animated films were only 42 per cent for those with only a grade-school education versus 63 per cent for high school graduates.) The superiority of the animated films tended, however, to hold for both above-average and below-average men, in terms of either educational level or mental-ability index (AQE score).

Effects of Different Animation Techniques on Learning of Nomenclature

In addition to comparing the effects of animated versus non-animated examples on proficiency at reading micrometer settings, effects of two different presentations (used in an introductory section of the films) were compared as to their relative effectiveness in teaching the nomenclature of the micrometer. One of the two kinds of presentation compared (Technique A) employed simply pop-in arrows, with no labels, to identify the parts of the micrometer as they were orally named. Technique B employed more complex animation, with camera effects such as "dissolves," and the labels were faded on and off the screen as the narrator described each part. (The narration was identical for Techniques A and B.)

Test results on the comparative effectiveness of these two techniques showed that Technique B (using labels, etc.) was on the average significantly superior to Technique A (simple, no labels). The average percentage of parts correctly named by the groups of men taught by Technique A was only 16 per cent; the corresponding

TABLE 2

Effects of Animation on Proficiency in Micrometer Reading as a Function of Audience Characteristics (Educational Level and A.Q.E. Score)

A. A.Q.E. Score	Animation		Non-Animation		Difference	
Highest (80-100)	11.55	10.81	11.76	9.90	(-.21)	(.91)*
Medium High (60-79)	10.07		8.05		(2.02)	
Medium Low (40-59)	6.87	6.02	5.65	4.24	(1.22)	(1.78)**
Lowest (00-39)	5.16		2.82		(2.34)	

B. Educational Level (Years of school completed)	Animation	Non-Animation	Difference***
8 years or less	6.30	4.39	(1.91)
9 years	6.11	4.88	(1.23)
10 years	7.10	6.05	(1.05)
11 years	7.48	6.70	(0.78)
12 years or more	9.42	7.56	(1.86)

* Superiority of animated over non-animated films for high AQE consistent for 13 out of 16 comparisons between paired class means; P-level: by sign test, .01; by Wilcoxon summed-ranks test for paired replicates: <.005.

** Superiority of animated over non-animated films for low AQE consistent for 12 out of 16 class comparisons; P-level: by sign test, .05; by summed-ranks test, <.005.

*** The suggested curvilinear trend of these differences requires further study, including similar analyses of additional data, before dependable conclusions can be drawn.

score for Technique B was 29 per cent. The margin of superiority for Technique B was thus a score gain of 13 percentage points.

Further analysis showed that the more elaborate Technique B—employing labels and camera "dissolves"—was advantageous primarily for terms not recurring frequently in later portions of the film. Other terms, used repeatedly in other connections, were evidently conveyed well enough by oral repetition so that for these terms the more elaborate labeling and animation technique showed less advantage. Thus the average percentages of fully correct answers for terms occurring frequently in other parts of film (mentioned 16 to 19 times in all) were 22.7 per cent for Technique A and 27.2 per cent for Technique B, giving a difference (advantage of B over A) of only 4.5 per cent; but, by comparison, the results for terms occuring infrequently (only two or three times in the film)—and hence dependent on the introductory nomenclature section for getting them across—were 7.0 per cent for A and 32.8 per cent for B, with a much larger advantage (25.8 per cent) for B over A (cf. Table 3).

This pattern of results was, like the findings for effects of the films on micrometer-reading ability, consistent for groups of trainees differing in intelligence and educational level. It was also consistent regardless of whether knowledge of nomenclature was strictly scored (correct spelling of terms) or loosely scored (credit given for a reasonably close approximation to the correct name), as is seen in Table 3. This suggests that the technique of using labels helped fix the names of the parts in the men's memories, in addition to giving them the correct spelling of the terms.

TABLE 3

Comparative Effects of Techniques A and B (See Text) On Items Testing Knowledge of Nomenclature, for Strict and More Liberal Scoring of Test Items

	Strict Scoring			Liberal Scoring		
	Percent Correct			Percent Correct		
	Tech. A	Tech. B	(Diff.) (B-A)	Tech. A	Tech. B	(Diff.) (B-A)
Frequent Terms						
Spindle	21.31	27.92	(6.61)	23.71	32.32	(8.61)
Barrel	22.11	23.82	(1.71)	28.20	29.43	(1.23)
Thimble	24.67	29.74	(5.07)	29.48	33.38	(3.90)
Average	22.70	27.16	(4.46)	27.13	31.71	(4.58)
Infrequent Terms						
Anvil	10.73	47.19	(36.46)	12.33	52.65	(40.32)
Frame	3.36	18.51	(15.15)	3.68	19.57	(15.89)
Average	7.05	32.85	(25.80)	8.01	36.11	(28.10)

Discussion

As anticipated, a material improvement in learning was effected by the use of the simple animation techniques employed. Several points concerning practical implications of this finding deserve discussion.

"Movie" Techniques for Teaching "Static" Material

The training objective of the experimental films was the teaching of what may be termed a non-moving or "static" subject matter. The essential skill to be taught did not depend primarily on perception of motion. The task was to read correctly a fixed instrument display (micrometer setting). Nevertheless, the use of animation techniques, characteristic of the instructional motion picture, was effective in increasing the training value of the instruction. This outcome raises a question concerning the frequently expressed opinion that, for so-called "static" subject matter, the choice of moving-picture film as the training medium is not warranted —that a static presentation (e.g., the typical film strip) is necessarily as good for presenting "static" material. The findings of the present study obviously do not tell the whole story, but they do furnish an example of improvement in teaching a static subject (reading fixed micrometer settings) resulting from the utilization of a characteristic motion-picture technique (overlay animation).

Alternative Methods

However, it is important to consider the probable reasons for the superiority of the animated films. Here we should distinguish clearly between underlying psychological or pedagogical factors and the production techniques employed for implementation. A good assumption is that much of the superior effectiveness of the kind of animation techniques studied here lies in the way they stress and clarify specific informational content, and in the way they direct attention to key aspects of a complex visual scene by precise timing of visual indicators keyed to the narration. It is worth noting that these stressing and attention-directing functions can also be achieved in other ways. Specifically:

1. In designing a motion picture presentation, other cinematic techniques such as live-action pointing, indicating, and writing—using suitable models, mock-ups, or other props—might serve the same attention-getting and stressing functions more cheaply, and as well or even better than in the present instance by overlay animation.

2. A logically important alternative technique not included in the present experimental work is the free use of arrows, labels, etc. in a static (non-moving) visual presentation such as a film-strip or series of slides. However, within the limitations imposed by usual production practices and commonly used still-projection equipment (a limited number of frames and relatively slow rate of changing frames) this alternative would generally be restricted to simultaneous rather than rapidly changing pointing to (and labeling of) the successively identified display elements.

3. Using static visual materials—films, strips, slides, etc.,—a first-rate classroom instructor might, using a pointer, suitable "chalk talk," etc., achieve the same instructional effects as the animated films here used—thus serving the stressing and attention-direction functions by classroom utilization rather than by film production. A good instructor <u>might</u> do this. Whether the average instructor generally <u>will</u> in fact do it, and do it as effectively as the film can, seems doubtful. There is one important advantage in programming the pointing-and-stressing devices into the film, by animation or equivalent technique: namely, effective implementation of these instructional functions can be effected once and for all, in advance—in standardized, well-timed fashion—by incorporating them into the film program itself.

Amount of Improvement Achieved

The problem to which this study was addressed, as stated at the outset, was: "Do trainees actually learn more through the use of motion-picture animation techniques?" Granted an affirmative answer, within the limits of the specific techniques compared in this study, the further question—how much more they learn, and how the amount of improvement can be assessed—can be dealt with. Is it, in particular, worth the added cost of providing the animation?

Present research techniques obviously do not provide an absolute, hard-and-fast answer to these latter questions. [Cf., however, Edwards (1956), Lumsdaine (1962a), and other approahces to the estimation problems involved here.] An absolute standard of comparison is lacking. For one thing, as noted earlier, the results can be expressed in various ways, yielding somewhat different comparative figures. Thus, the averages and percentages themselves have only a relative significance, not an absolute one. We can, however, provide a very rough yardstick in terms of comparative instructional potency of other elements in the film-training situation. The illustrative data presented in Table 6 (see Section II), for films used without supplementary slide-film material, are instructive in this respect. From the tabulation in Table 6-A it can be seen that the advantage gained by animating the three initial examples of micrometer reading (used in the "baseline" film) was appreciably less than that achieved by the presumably less expensive device of adding three more examples (which required, however, two extra minutes of class time.) These comparative results give some basis for evaluating the improvement made by animation. It might further be inferred from the data that animation of the instructional material would be unnecessary, provided time is allowed for additional examples. This conclusion does not necessarily follow. In the present instance, use of animation in the examples produced an added gain, above and beyond that attained by the maximum amount of repetition. (This can be seen from the data presented in Table 6-B: Even the longest film treatment was still further improved by animation techniques.)

Improvement in Teaching Nomenclature Produced by Animation Employing Visual Labels

In teaching nomenclature, Technique B was found to be superior to Technique A. These techniques differed in two respects: (1) Technique B used printed labels giving the names of each part visually as it was named by the narrator, whereas Technique A did not use these labels; and (2) Technique B "dissolved in" each part as it was named, whereas in Technique A all parts of the micrometer were present throughout the presentation of nomenclature, with each part identified, as it was named, by a pop-in arrow. Since the two techniques differed simultaneously with respect to two factors—labels and camera dissolves—the present data do not establish which factor was responsible for the superiority of Technique B. However, it seems reasonable to suppose that the use of labels to strengthen the verbal identification of the parts was probably a critical factor. (Cf., the effects of narration, Chapter 13.) Until more data are at hand it seems plausible to assume that free use of visual labels in pictorial film material will markedly improve the learning of nomenclature, particularly when insufficient opportunity is given for learning the names of the parts in other sections of the film.

"Confounding" of Micrometer-Reading and Nomenclature Variations

Finally, it may be noted that the films using animated examples for teaching micrometer reading uniformly employed Technique B for teaching nomenclature, while the films with non-animated examples used Technique A. Although conceivably there could have been some interaction between the techniques used for teaching nomenclature and micrometer reading, this seemed so unlikely that it was decided to confound the two experimental factors for the sake of economy. Relevant to this presumption of substantial functional independence of the two factors is the fact that a material advantage for Technique B was found only for terms ("frame," "anvil") that did not enter into the task of reading the micrometer settings.

Summary

The training value of simple animation techniques as cues in teaching an instrument-reading skill was studied by a controlled field experiment. The study was performed by testing 32 classes of Air Force basic trainees to find out how well they could learn to read micrometer settings from seeing one of several specially prepared experimental training films on the micrometer. The films seen by half of the men employed simple animation devices—pop-in labels, moving arrows, etc.—superimposed over the pictorial material. These animation devices were designed to stress and direct attention to key aspects of the pictures. The films seen by the other half of the men were identical except for the omission of these animation devices in key sequences of the films. Test data for approximately 1,300 trainees showed that use of the animation cue devices consistently produced a marked increase in the amount men learned from the film, thus lending factual support to the practice of using such devices in technical films.

Section II. The Effect of Repeated Examples

As McGuire has noted in Section I of Chapter 13 (q.v.), it is commonly believed that learning tends to increase with repetition. When a skill such as instrument-reading is taught in a training film by means of demonstrational examples, repetition can be employed in various ways. For instance, the film may be shown over again in its entirety. But a considerable portion of the footage in many films consists of time-consuming sequences such as credit-titles, introductory explanation, and other materials for which repeated presentation may be unnecessary and boring. (The same objection, among others, arises in the re-cycling of lesson materials in teaching machines—a feature abandoned by Skinner and others in recent designs. Cf., Lumsdaine 1959c, p. 12.)

It seems preferable, therefore, to illustrate repeatedly only the most pertinent instructional material. This can be done either by incorporating additional examples into the film or by adding supplementary material. With either of these methods there is the added potential advantage that more varied examples can be shown than are afforded by merely repeating the film in its entirety.

In the present study of the use of multiple examples, evidence was obtained on the following questions: (1) To what extent is the training value of a film program affected by increasing the number of instructional examples in the film? (2) Similarly, how is the amount of learning affected by using additional examples incorporated as supplementary material at the end of the film? (3) Is a "saturation point" soon reached after which no more gain is produced by the addition of further examples? (4) Can common kinds of errors made by the trainees be reduced by designing additional examples to take specific account of the tendency to make these errors? (5) For what kind of material is the use of additional examples most helpful? Finally (6), who benefits most from being given additional examples?

Method

Experimental Film Materials

Multiple Examples Within the Film. Two experimental versions of both the animated and non-animated films described in Section I were used in this study. One version contained only the three demonstrations of micrometer-scale readings from the original film, while the other included three additional, specially prepared examples. Each film opened with shots of several types of micrometers, leading into a sequence on the nomenclature and principle of the micrometer. The method of obtaining readings was then demonstrated by a series of examples. Each example showed how to read a different setting, but the development of the examples followed a fairly constant pattern, illustrated by the excerpt that was presented in Section I. The narration for another example—the first given in the film—read, in full, as follows (with pauses indicated by the dots to allow time for student to perceive the accompanying visual depiction).

> "The measurement is read by first reading the number fully visible on the barrel. This is 6. Therefore, it is 600 one-thousandths... Beyond the 6 mark is visible another graduation. This gives us another 25 thousandths...On the thimble scale there is shown the graduation that has stopped on the index line—the one-thousandths graduation ... The full reading is .626."

Following this reading, additional examples were interspersed among sequences on various topics such as care and handling of the micrometer. Table 4 shows the content of the two different film versions. It should be noted that the three-example and six-example versions were identical except for the inclusion or omission of the three extra examples.

Additional Examples Given by Supplementary Instruction. Half of the trainees who saw the shorter film version, and half of those who saw the longer version, were also given a four-minute supplementary slide-film with a recorded commentary. This supplementary instruction concentrated on how to read difficult settings which fall just short of a division on the main (barrel) scale. Preliminary experiments had shown that these were stumbling blocks that accounted for an unduly large share of error. The supplementary instruction was designed to reduce such errors by

TABLE 4

Three-Example Version (11 min)	Six-Example Version (13 min)
Example No. 1: Reading a micrometer setting of .626	Example No. 1: Reading a micrometer setting of .626
Sequence on available sizes of micrometers	Sequence on available sizes of micrometers
	Example No. 2: Reading a micrometer setting of .532
Sequence on "feel" of taking measurements	Sequence on "feel" of taking measurements
	Example No. 3: Reading a micrometer setting of .015
Sequence on the abuses of the "mike"	Sequence on the abuses of the "mike"
Example No. 2: Reading a micrometer setting of .751	Example No. 4: Reading a micrometer setting of .751
More on "feel" of taking measurements, etc.	More on "feel" of taking measurements, etc.
Example No. 3: Reading a micrometer setting of .750	Example No. 5: Reading a micrometer setting of .750
Sequence on inside micrometers, depth-micrometers, etc.	Sequence on inside micrometers, depth-micrometers, etc.
	Example No. 6: Reading a micrometer setting of .498
More on care of micrometers, etc.	More on care of micrometers, etc.

discriminating settings "just above" versus "just below" a main scale graduation. Three examples of this type were given (e.g., 0.597), plus one additional example of the easier type of setting.

The micrometer settings used as instructional examples in the basic (three-example) films were .626, .750 and .751. The six-example films used these examples and also .532, .015, and .498. The length of the films ranged from 11 minutes (for the three-example films) to 13 minutes (for the six-example films), and the supplementary slide film used for half of the class groups required six minutes. The lengths of corresponding animated and non-animated films were, of course, identical.

Experimental Design and Procedure

Basic features of design and procedure have already been described in Section I. (See also Supplementary Table 2 at the end of the chapter.) It will be noted that thirty-two class groups of trainees were tested, totaling about 1,300 men, with each of the class groups consisting either of one flight (50 to 60 men) or of a half-flight (25 to 30 men). Sixteen of these classes saw the three-example film. The other sixteen classes saw the six-example film. For half of the classes in both groups the film was followed by the supplementary instruction, containing four additional examples. Thus eight classes saw the three-example film only, eight classes saw the six-example film only, eight classes saw the three-example film plus the supplementary instruction, and eight classes saw the six-example film plus the supplementary instruction. It should also be noted that half of the classes in each subgroup (16 classes in all) were given the pre-film test before seeing their film, while in the other 16 classes this pre-test was omitted. All assignments of classes to treatments were purely random.

Results

Overall Effects of Repetition

Effect of Additional Examples Within the Film. Data on initial ability and average gains produced by the films, as well as information on data analysis, have already been described in Section I. It will be recalled that, because of the very low level of pre-film knowledge, knowledge shown after the film is almost entirely attributable to the effects of the film instruction. With a greater number of opportunities to observe micrometer settings within the film, the amount of knowledge increased materially. Comparison of average results for all groups seeing the three-example film with those for all groups seeing the six-example film, in terms of the average score made on the post-film test, showed a mean of 6.48 for the shorter film versus a mean of 8.46 for the longer film. With the 15-item test used, since the mean test score reached after viewing the basic three-example film was about 6.5, the total room for demonstrable improvement in scores was about 8.5 (that is, 15.0 less 6.5). The average gain of about two points produced by the three additional examples may thus be said to represent about one-fourth of the maximum possible improvement.

Reliability of the Findings. As in Section I, extensive checks were made to insure that these findings were not due merely to chance or to special, unusual circumstances. In terms of statistical probabilities (see analysis of variance in Supplementary Table 3) there is less than one chance in a thousand that a difference in performance as great as that shown above would arise solely through chance factors. The reliability of the results is again attested, also, by the fact that the same pattern of results was consistently obtained under a variety of experimental conditions. These statements apply also to the overall results on effects of supplementary instruction, presented below.

Gains Produced by Supplementary Slide-Film Instruction. As anticipated, those men who received supplementary instruction learned considerably more than those who did not. The average scores were 8.68 with supplementary instruction and 6.30 without. The supplementary instruction was effective after either the shorter (three-example) film or the longer (six-example) film. But the effectiveness of giving further examples by supplementary instruction changed with the number of examples already seen in the film itself. When there were only three examples in the film, the contribution of the supplementary instruction was relatively great. The increase was much smaller when the supplementary instruction followed the longer six-example film. After the three-example film, an average of 4.6 questions were answered correctly without supplementary instruction, versus an average of 8.4 questions answered correctly with supplementary instruction. Thus, here the improvement due to the supplementary instruction was 3.8. After the six-example film, however, an average of 8.0 questions were answered correctly without supplementary instruction, versus an average of 9.0 questions answered correctly with supplementary instruction. Here the improvement due to the supplementary instruction was only 1.0.

The Trend of Improvement as More Examples are Added. As just seen, the supplementary instruction was less effective after the longer film. This implies a trend, with the amount of gain to be expected from adding further examples decreasing as the number already seen goes up. This trend is shown in Table 5. The figures in the second column indicate the amounts of further gain as more examples are added.

TABLE 5

Trend of Improvement with Increasing Number of Examples

	Number of Settings Correctly Read	
	Mean Score for Each Treatment	Increment from Additional Examples
Instruction with:		
3 Examples	4.57	--
6 or 7 Examples (averaged)	8.18	3.61
10 Examples	8.96	0.82

The diminishing rate of increase as further examples are added suggests the existence of a maximum number of demonstrations after which little or no further improvement would result from adding still further examples. The determination of this "saturation point" would require further experimentation with the number of examples still further increased beyond the maximum number used here. In the present instance it is of interest to note that adding still further examples produced a significant gain even after the number of examples in the original film had been doubled. The saturation point for any instructional program would be expected to depend on several factors, including the difficulty of the material and the intelligence of the audience. Some evidence on the relation between these factors and the gains resulting from use of additional examples will be presented later.

Effects of Repetition and Animation Compared. Table 6, which has been previously referred to in Section I in connection with the magnitude of gains produced by animation cues, compares the gains thus produced with those effected by adding examples. It can be noted in the present context that the gain produced by adding three or four examples to the basic non-animated, three-example film markedly exceeded that resulting from the probably costlier technique of adding animation (Table 6-A).

TABLE 6

Average Scores (Mean Per Cent of Correct Answers)
For Animation and Repetition Factors

A. Overall Scores

Average score for basic (3-example) non-animated film	25.0
Average score for basic (3-example) animated film	35.9
Gain by animating the examples in basic film	10.9
Average score for basic 3-example film (without animation)	25.0
Average score for film with three additional instructional examples (six examples in all), without animation	48.4
Gain by adding three non-animated examples	23.4

B. Sub-group Scores Showing Interaction of Animation and Repetition Factors

Average scores for:	Non-Animated	Animated	Gain from Animation
Three-example film	25.0	35.9	(10.9)
Six-example film	48.4	57.8	(9.4)
Film instruction with ten examples (6 in movie plus 4 in auxiliary slide-film)	52.1	67.3	(15.2)

Analysis of Results for Different Sub-groups and Kinds of Material

Comparative Advantage of Supplementary Instruction for Easy and Difficult Material. Analyses were made of data on individual test items, and sub-sets of test items, as well as on total scores. These were needed for examination of the results as dependent on difficulty of the material as well as to check consistency of the overall findings. Percentages of men giving partially correct and fully correct answers to each of the 15 items in the post-film test were tabulated for those who saw the basic (three-example) film or six-example film, with and without the supplementary slide-film. A special analysis was made of the effect of the supplementary instruction on the most difficult kinds of settings, which the basic film failed to teach men to read effectively. (These were the settings that fell just a few thousandths below a main scale division on the instruments.) Results showed that the supplementary slide-film, which concentrated on these difficult settings, had significantly more effect in reducing errors on this difficult material than it did on the easier materials. Since the base level attained by the basic film was much different for the easy and difficult settings, this comparison was made in terms of reduction in errors relative to amount of room for improvement for each type of setting (absolute error reduction divided by total number of errors made by those who did not get the supplementary instruction). In these terms, the advantage of film-plus-supplementary-instruction over film alone averaged only 21 per cent for the easier settings, as against 31 per cent for the difficult settings.

It is also interesting to note that it was the supplementary instruction that was particularly effective in reducing the characteristic type of mistake of reading the difficult settings too high by 25 thousandths of an inch. The supplementary slide-film pointed out the danger of making this error and emphasized how to avoid it. This emphasis evidently paid off. Without supplementary instruction the proportion of men making the characteristic error was about as high (23 per cent) as the proportion getting the correct answer on the difficult settings. But after getting the supplementary instruction 50 per cent read the difficult settings correctly and only about one-fourth this many (13 per cent) made this particular kind of error.

Comparison of Kinds of Added Examples Within Film versus In Supplementary Slide-Film. Here the emphasis is on the kind of examples used rather than on the fact that one set was included in the film and the other presented at the end by a supplementary slide-film. The latter distinction was a more or less accidental aspect of the experiment, and the two modes of presentation were quite similar, since most of the film material was relatively static, and since similar recorded narration was used to accompany both the inserted examples and the supplementary slide-film. The examples inserted in the film followed much the same pattern as the original three examples, giving two more "easy" settings and one "difficult" setting. This contrasts with the supplementary examples' concentration on avoiding errors on the difficult setting (by showing how to make the proper discrimination between easily confused readings).

As can be seen by comparing the overall gains already reported above, the average gains for the two sets of examples were roughly the same. The three within-film examples increased the gains over the basic film by about one-fourth of the maximum possible improvement; approximately the same was true for the supplementary slide-film examples. However, the average values are influenced by the proportion of easy and difficult settings included in the test. A more incisive comparison is thus afforded when the data are analyzed separately for these two types of material, as shown in Table 7.

These results show that, for the easier material, the two kinds of additional examples were about equally effective in improving the value of the basic film—about a 34 per cent gain versus 35 per cent. But for the difficult material, the

TABLE 7

Effects of the Within-Film and Supplementary Examples on Easier and More Difficult Material (Percent of Possible Improvement Produced)

	For Easier Settings	For Difficult Settings
Three added examples in the film:	33.9	29.0
Four supplementary examples:	34.6	42.4
Difference:	0.7	13.4

specially designed supplementary examples were considerably more effective (42 per cent gain) than were the less specialized (within-film) examples (29 per cent gain).

The Factor of Intelligence. Analysis of the results for men of differing educational levels did not show, in this case, a highly reliable pattern of results. There was a tendency for those with more schooling to benefit more from the added demonstrations, but the overall trend was not marked enough to be definitive. However, when trainees' scores were segregated according to intelligence test (AQE) scores, an interesting finding emerged: it was found that the more intelligent the trainees the greater was the benefit from the added examples. These results are illustrated in Table 8. The gains given in Table 8 are for the combined effect of all seven additional examples. Substantially the same pattern is shown when the results are calculated separately for effects of the film-presented examples and for effects of the examples in the supplementary slide-film. The differences between the highest and lowest AQE classes are in each case reliable at better than the .01 probability level.

Although it would be expected that the brighter men would learn more from the basic film, it may be surprising that relative benefits from additional examples were also much greater for the bright trainees. It is sometimes held that repetition is needed only for the duller men—that it is boring and needless for the brighter student. In this instance, on the contrary, the results show it was the brighter men who benefited most from additional demonstration.

Number of Examples Related to Intelligence and Difficulty of Material. Assuming it is desired to obtain maximum instructional returns from demonstrational instruction, how much repetition—how many examples—can profitably be used? In general, we would expect that continuing to give more and more examples would (a) be necessary to the extent that the material is difficult enough to require it, but would (b) pay off only to the extent that the audience is intelligent enough to profit from the further repetition. These relationships are suggested in the two charts

TABLE 8

Percent of Possible Gain Produced by Added Examples
For Men of Different Intelligence Levels

A.Q.E. Score	Percent of Maximum Gain Possible		
	10 ex. vs. 3 ex.	6 ex. vs. 3 ex.	Supp. vs. non-supp.
Highest (80-100)	83.3	61.9	64.1
High Average (60-79)	53.3	39.4	32.7
Low Average (40-59)	35.1	24.4	28.3
Lowest (39 or less)	12.0	9.8	2.3

comprising Figure 1, where the data on easier and harder material are plotted separately for scores achieved by brighter and less bright men after varying amounts of demonstration.

In Figure 1 it is seen that a limited amount of repetition—up to six or seven examples—was of substantial aid regardless of intelligence or difficulty of material. This is shown by the uniformly significant increases in score as the number of examples demonstrated is increased from three examples to six or seven examples. Beyond this point, the effectiveness of still further demonstration varied both with intelligence and with difficulty of the material. For the easy material (chart on left) only a limited amount of further demonstration seemed to be profitable; past six or seven examples, no further improvement is shown for the less intelligent and very little (not enough to be reliable statistically) for the more intelligent. But for the difficult material (right-hand chart) the data indicate that further gains are found by adding still further demonstration (up to 10 examples), particularly for the more

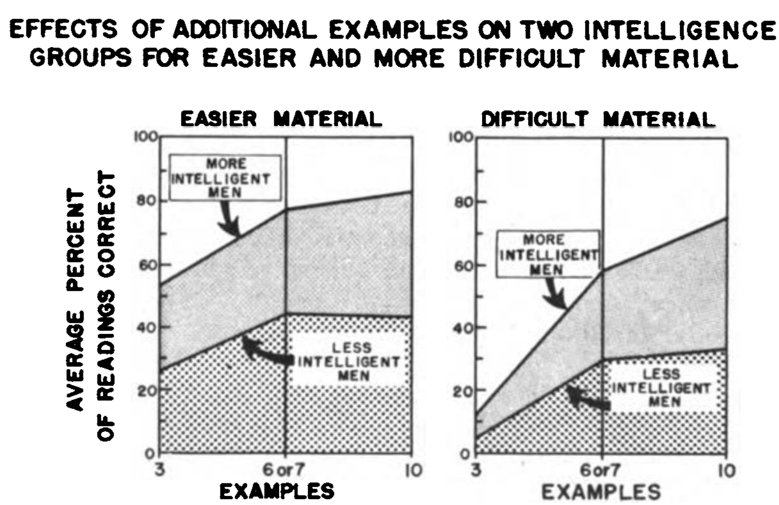


Figure 1. Gains from repetition as a function of difficulty and of intelligence of subjects

intelligent men. (The further gain for the more intelligent men is highly reliable; also they gained significantly more from the last increment in number of examples than did the less intelligent.)

Discussion

Multiple Examples as a Form of "Repetition"

The use of multiple demonstration examples is one obvious way to obtain the advantages of repetition in an instructional program. The experimental results just described make it clear that where this form of repetition is appropriate it can produce very substantial gains in the training value of a film. Some ways of gauging its importance in the present instance are suggested immediately below. Other implications of the findings, discussed later, deal with questions of the optimum amount and kind of repetition that should be used in instructional media.

Importance of Repetition in Determining the Training Value of a Film

It is clear from the findings that the repetition of demonstration in the form of further examples was a fairly potent factor. Adding only three additional examples produced roughly twice as much gain as did the use of overlay animation techniques employed in the original film. (See Table 6.) Another way to gauge roughly the importance of the factor of repetition is to compare its influence with differences in performance due to the highly potent factor of variations in intelligence among members of the film audience. The most intelligent major sub-group of the audience (those with AQE scores of 80 or better, comprising about a fifth of the men) averaged about 11-1/2 correct answers out of 15 test items, while the least intelligent group (AQE below 40, comprising about a fourth of the men) answered an average of only four items correctly. This means an effective range of about 7-1/2 points between these "brightest" and "least bright" groups. The average gain from adding only three examples to the basic film was about 3.4 points, which is over 40 per cent of the average difference between these "brightest" and "least bright" groups. The average gain from all seven additional examples was 4.4 points, or nearly 60 per cent of the difference associated with the high versus low intelligence groups.

There are not at present enough data on the comparative influence of other elements in training-film instruction to assess fully these indices of the relative effectiveness of the factor of repetition. However, on the basis of the comparisons just made, the speculation can be hazarded that the homely technique of using numerous illustrative examples may turn out to be one of the more potent determiners of demonstrational teaching.

How Much Repetition is Enough?

When repetition is one of the techniques employed in order to increase instructional gains from a film, the question arises: How much repetition should be employed? The answer evidently involves determining a "saturation point," after which more repetition will fail to achieve further gains. In a study of repetition by repeated re-showing of an entire film, reported by McTavish (1949), it was found that little or no further gain was produced after one repeat: "...the contribution made by repetition of showings fell off rapidly after the first repetition...only the increment attributable to the first repetition is statistically significant." McTavish's study and the present one thus both indicate the existence of a "saturation point," although one study dealt with repetition of entire films and the other study with repeated demonstration within a film. The important thing, however, is not the mere existence of a "saturation point;" rather, it is the question of what determines where this point will occur.

It would be unfortunate if such results as those of McTavish led to the erroneous notion that there is a more or less fixed, low value for the number of repetitions that can be usefully employed. How much repetition should be used in any instructional program undoubtedly depends on a number of factors, some of them as yet not identified. The results of the present study underline three important points: first, the number of useful repetitions may be considerably more than one (this could well be the case for whole-film repetition, under suitable conditions, as well as for intra-film repetition); second, the amount of repeated illustration that will be helpful is not a constant, but depends on characteristics of the audience and the film material; and third, some of these characteristics, at least, can be identified and their relation to the location of a "saturation point" can be determined. The present evidence points to two such factors—the difficulty of the material and intelligence level of the audience.

It is not surprising to find that the more difficult material will stand more repetition before reaching a saturation point. The relation between intelligence and useful amount of repetition is perhaps less obvious. It might be thought that repetition would help only for the less intelligent, that the more intelligent would not profit by it. From this standpoint it would appear surprising that it was, on the contrary, the more intelligent who benefited most from further repetition.

However, there is no real contradiction between the experimental findings and the expectation of more need for repetition among the less intelligent. The apparent contradiction disappears as soon as a distinction is made between <u>need</u> for repetition and ability to profit from it. The present data show that the less intelligent men <u>do</u> need more repetition than do the brighter men in order to reach a certain minimum level of accomplishment—say 40 per cent correct readings. But at the same time the gains from a given amount of repetition are greater for the brighter men. The latter do not need as much repetition to reach a minimum level of performance—but they are better able to profit from a reasonable amount of further repetition, especially on the difficult material.

In general, taking such factors as intelligence and difficulty of material into account should provide some guidance in determining how much repetition to employ, although at present an exact answer for any particular subject matter and kind of students would have to be determined by specific experimentation (see next section). From the present data, however, it seems a good guess that the useful limit on amount of repetition often may be considerably higher than the amount currently used in many training films. The prevailing tendency among film makers, and probably others who prepare instructional materials, is probably to make too little use of repeated illustration, rather than too much. In cases where doubt exists, the decision to illustrate more fully or use more repetition than is thought necessary would very likely increase the training value of many instructional programs.

What Kind of "Repetition" is Most Useful?

One question to be asked here concerns the relative advantage of "varied" versus "identical" repetition—e.g., using varied examples versus going through the same examples more than once. Further research on the various techniques for employing repetition is needed. The films used in the present study employed "varied" demonstration exclusively, and hence afford no direct evidence on the question. (See research note in the Appendix on the study by Kanner and McClure, 1956.) A second question relates to determining which aspects of the material should be stressed in providing repetition of a procedure by additional demonstration. Here the present evidence suggests that identifying and concentrating on the more difficult aspects of the materials (after a few initial easier examples) is a profitable procedure. [Cf., the analogous finding by Kimble and Wulff, Chapter 15, on the special value of "response guidance" for more difficult aspects of subject matter.] In the present study this was done by the supplementary slide-film, which concentrated on a single and relatively infrequent type of difficult setting.

Pre-production Testing

It is important to note that one particular type of micrometer setting was a special source of difficulty. This was only apparent after preliminary research testing had been done. The implication here is that, in order to be able to concentrate repetition in an instructional program on the really difficult aspects, preliminary testing (before the final script is written) may be required in order to pin-point the aspects of the subject that are the main stumbling blocks. This "pre-production research" should go beyond the usual film-script research that is concerned with only the technical accuracy of the presentation; it should include experimental teaching and testing of the subject matter with a representative sample of the intended audience. In some instances this can be done by using a preliminary filmstrip mockup of an intended film, in other instances by "live demonstration" teaching with suitable props and mockups.

Such testing of a preliminary version of the instructional sequences to be filmed would be needed to get an exact indication of the optimum amount of repetition for a particular film. In the absence of this kind of testing, an analysis of the training task in terms of such factors as the difficulty of material and the intelligence of the audience can perhaps furnish a useful advance guide—e.g., use most repetition on the more difficult parts of the subject matter, less on the easier parts. But the exact amount of repetition needed for any subject matter can, at the present state of the art, be determined only by pre-production testing with the particular subject.

Building Repetition Into a Film Versus Alternative Procedures

The advantages of repeated examples built into the experimental films used in this study were also obtained with supplementary materials designed to present additional examples. If these materials are as well designed and carefully prepared as the film itself they can do a good job and lessen the need for repetition within the film.

However, good supplementary materials are not always readily available, and in practice are often poorly utilized. The supplementary materials used in the present experiments were unusual in that they were virtually a part of the film. They employed similar pictorial materials automatically projected with pre-determined timing and sequence, geared to an optically recorded sound track that formed a continuation of the narration of the film. It is probably rare for supplementary materials to be as carefully prepared and used in practical classroom instruction situations; consequently the present results definitely do not demonstrate that attempts to provide repetition by classroom utilization of supplementary materials will be as satisfactory in practice as providing repetition within the film itself. (In the present experiment, use of slides rather than further film footage to present the further additional examples was adopted largely as a matter of expediency, and, as previously noted, the main interest in the "within-film" versus "supplementary" comparison was in the kind of examples used, not in the fact that one was presented on 16mm film and the other on 35mm slides.) An important advantage of building additional examples or other forms of repetition into the film itself is that satisfactory implementation of the factor of repetition can be determined and effected in advance, once and for all, in standardized, well-timed fashion.

Cost Factors

The incorporation of more repetition into films will generally result in somewhat longer and sometimes more expensive films. Where added expense is involved, however, the extra cost may be more than offset by the increased training value of the film. But added footage devoted to repetition will not always necessarily increase film budgets, since the cost of this kind of footage may be relatively slight as compared with that of other, more expensive techniques. Accordingly, if cost must be kept to a minimum it may sometimes be better to economize on other factors than to cut down on use of repetition.

Summary and Implications

This part of the study was designed to investigate the instructional value of multiple examples as in the films. The films seen by half of the men contained three different examples of micrometer reading, while experimental films seen by the other half contained six different examples. Following the film, half of all groups were given supplementary instruction with four additional examples. This supplementary instruction was given by a slide-film with a recorded commentary, and dealt primarily with the more difficult kinds of micrometer settings. Results showed that the amount learned increased consistently with a greater number of examples, whether the examples were given in the film itself or in the supplementary instruction. Giving further supplementary examples made a significant improvement even after the number of examples in the original film had been doubled, but the rate of improvement diminished as the number of examples increased, suggesting a "saturation point," dependent on the difficulty of the material, after which further examples would fail to produce more gain. Common mistakes were made less often when supplementary instruction was used to emphasize the difficult material, and the more intelligent men benefited even more from additional examples than did the less intelligent men.

The following recommendations can be offered:

1. For producers or programmers of instructional media. To increase the training value of instructional programs (e.g., films and film-strips) designed to teach technical information and skills:

a. Make extensive use of repetition—especially, where appropriate, varied repetition in the form of numerous demonstrational examples. (Use of even more illustration than seems necessary may well increase the amount learned from a film.)

b. To help insure a uniformly high level of instructional effectiveness, build needed repetition into the film (or other programmed material) itself rather than relying on the use of supplementary aids.

c. Do not dispense with repetition in special materials intended for selected, high-intelligence audiences. (The more intelligent may be able to profit from repetition even more than do less intelligent men.)

d. Concentrate the use of repeated illustration on those aspects of the material that are most difficult for the trainee to master (assuming that these represent points that it is important to get across).

e. To identify difficult aspects and to determine how much illustration or repetition is needed, try to obtain detailed pre-production test data from typical trainees before and after exposure to classroom instruction which parallels that proposed for the film or other instructional program.

2. For teachers. To maximize the training value of those existing films, TV kinescopes, or other instructional programs in which insufficient repetition has been employed:

a. Present the program twice or more, depending on the difficulty of the material, or if feasible repeat key sections of it.

b. Obtain and use supplementary materials—e.g., filmstrips, charts, models, etc.—to provide additional examples or other appropriate forms of varied repetition.

c. In providing repetition in classroom instruction, apply the considerations noted in 1-c, 1-d, and 1-e above to determine how much repetition to use and on which aspects of the subject matter it should be concentrated.

SUPPLEMENTARY TABLE 1

Mean Percentages, For All Classes That Saw Animated Films and For Those That Saw Non-Animated Films, of Partially and Fully Correct Answers to Each of 15 Post-Film Test Items on Micrometer Reading

Test Item		(a) Percentages giving fully correct answers			(b) Percentages giving at least partially correct answers*		
Micrometer Setting	Item No.	Anim.	Non-An.	(Diff.)	Anim.	Non-An.	(Diff.)
.022	1	51.9	53.4	(-1.5)	58.6	65.7	(-7.1)
.012	7	64.5	58.3	(6.2)	72.7	75.5	(-2.8)
.425	6	66.8	54.6	(12.2)	71.3	65.8	(5.5)
.206	10	63.0	46.8	(16.2)	65.9	53.7	(12.2)
.575	15	57.4	46.5	(10.9)	62.8	57.5	(5.3)
.537	2	51.6	39.9	(11.7)	53.4	44.6	(8.8)
.731	8	64.6	45.8	(18.8)	67.1	52.6	(14.5)
.114	9	61.3	47.6	(13.7)	65.1	58.3	(6.8)
.255	11	63.3	47.4	(15.9)	67.0	54.6	(12.4)
.182	14	45.4	35.1	(10.1)	46.9	39.9	(7.0)
.501	12	54.6	42.5	(12.1)	57.5	47.9	(9.6)
.802	13	58.3	44.2	(14.1)	60.6	50.5	(10.1)
.697	3	34.1	32.5	(1.6)	35.2	36.7	(-1.5)
.471	4	44.8	37.8	(7.0)	46.4	42.6	(3.8)
.848	5	28.1	29.8	(-1.7)	29.3	32.2	(-2.9)

*Answers given credit as being at least partially correct included (in addition to entirely correct answers) those answers which were incorrect in one or more of the following respects, but otherwise correct: decimal point was omitted, or misplaced; an extra decimal point was inserted; one or more extra zeros were inserted before the correct digits; a zero was omitted before the correct digits; one or more zeros were added after the correct digits; or a zero was omitted after the correct digits.

SUPPLEMENTARY TABLE 2

Class Means (Mean Number of Micrometer Settings Correctly Read After Exposure to Instruction) for Each of the 32 Classes Tested, for Three- and Six-Example Film Groups, With and Without Supplementary Slide-Film

Experimental Conditions*		First Replication				Second Replication			
		No Supplement		With Supplement		No Supplement		With Supplement	
P.F.T.	An.	3-Ex.	6-Ex.	3-Ex.	6-Ex.	3-Ex.	6-Ex.	3-Ex.	6-Ex.
A. Overall Values (Total N=1283)									
+	+	6.633	9.464	10.769	10.759	5.897	9.127	9.073	10.891
+	-	4.172	8.893	8.115	7.042	3.857	7.340	10.059	7.217
-	+	4.276	7.586	7.792	10.276	4.724	8.500	6.849	8.472
-	-	4.136	6.682	9.552	8.240	2.857	6.157	4.940	8.768
B. High AQE** (Total N=541)									
+	+	7.50	10.44	13.61	13.18	7.83	11.75	10.58	13.52
+	-	5.08	10.83	9.18	11.67	6.48	10.75	13.23	10.18
-	+	7.27	9.91	14.40	12.86	8.14	11.76	9.95	11.44
-	-	6.40	9.00	12.50	13.30	5.04	9.94	8.29	11.19
C. Lower AQE** (Total N=726)									
+	+	5.33	7.70	8.42	7.33	5.03	7.47	7.70	9.00
+	-	3.53	5.40	7.33	4.27	2.35	4.52	7.66	4.50
-	+	2.53	5.31	6.05	7.36	2.64	6.71	5.39	5.82
-	-	3.47	5.08	6.80	4.87	1.38	4.09	3.69	6.52

* "P.F.T." indicates presence (+) or absence (-) of pre-film test given prior to instruction; "An." indicates presence (+) or absence (-) of overlay animation used in the within-film examples.

**"High" AQE: men with AQE scores of 60 or over; Lower AQE group: men with scores below 60. Total N for parts B and C of Table 1 is 1,267; for the remaining 16 men of the total 1,283 no AQE scores were available.

SUPPLEMENTARY TABLE 3

Summary Statistics for Analysis of Variance, t-Test and Non-Parametric Significance Tests, Computed From Basic Data Summarized in Preceding Tables

A. Analysis of Variance for overall performance on micrometer reading (Based on 32 class means)

Source of Variance	Sum of Squares	d/f	Variance
(1) Animation	16.6182	1	16.62*
(2) Use of pre-film test	11.8850	1	11.88*
(3) Repetition (within film)	31.4258	1	31.43*
(4) Repetition (in supplementary slide-film)	46.3463	1	46.35*
(5) 1st order interaction of (3) & (4)	16.0753	1	16.08*
(6) Higher-order interactions involving (3) x (4), combined	7.0378	3	2.35
(7) All other 1st and higher order interactions, combined	6.4925	7	0.93
(8) Replications variance of class means within cells	19.2550	16	1.20
Total	155.1359	31	

B. Analysis of Variance for overall performance on nomenclature (Based on 32 class means)

Source of Variance	Sum of Squares	d/f	Variance
(1) Animation Techniques	3.6773	1	3.68**
(2) Other factors (repetition, pre-test), combined	0.1256	3	0.04
(3) All 1st and higher order interactions, combined	1.0361	11	0.09
(4) Replication variance of class means within cells	0.3120	16	0.02
Total	5.1510	31	

* $p < .01$, using within-cell variance as error estimate. Other tests, including non-parametric tests were also made; e.g., for animation, with $t = 3.75$ (15 d/f) with observations paired on basis of showing conditions and replication, where $p < .002$; Sign test: $p = .01$; Wilcoxon summed-rank test for paired replicates: $p < .005$.

** $F = 184$ ($< .001$); t for paired observations (15 df); p for sign test and summed-ranks test also significant.

SUPPLEMENTARY TABLE 4

Mean Scores on Micrometer-Reading Test After Seeing Animated Versus Non-Animated Films, For Subgroups Tested Under Various Instructional Conditions

	Mean post-film scores		
	Anim.	Non-Anim.	(Diff.)
A. Use of pre-film test:			
Subgroup given pre-film test	9.08	7.09	(1.99)
Subgroup not given pre-film test	7.31	6.42	(0.89)
B. Classroom conditions:*			
1. Morning Classes	9.34	7.80	(1.54)
2. Moderately warm afternoons	5.53	4.40	(1.13)
3. Hot afternoons	7.62	6.16	(1.46)

C. Distance men sat from screen:

	First Replication				Second Replication		
	Anim.	Non-an.	(Diff.)		Anim.	Non-an.	(Diff.)
Close: First 2 rows	8.40	6.62	(1.78)	Close: 1st 3 rows	7.94	6.84	(1.10)
Med.: Next 2 rows	7.88	6.67	(1.21)	Med.: Next 4 rows	8.23	6.87	(1.36)
Far: Next 2 rows	9.30	6.37	(2.93)	Far: Next 4 rows	7.11	5.16	(1.95)

* This comparison is necessarily confounded to some extent with other factors in the experimental situation. Means for condition 1 (morning) are based on nine pairs of classes, for condition 2 on three pairs of classes, and for condition 3 on four pairs of classes.

CHAPTER 18

FACTORS INFLUENCING THE EFFECTS OF STUDENT PARTICIPATION ON VERBAL LEARNING FROM FILMS: MOTIVATING VERSUS PRACTICE EFFECTS, "FEEDBACK," AND OVERT VERSUS COVERT RESPONDING[1]

Donald N. Michael
Nathan Maccoby

Introduction

Previous research has demonstrated that learning from a factual film can be increased with the use of audience-participation procedures. Audience participation in these experiments consists of having the students (or "audience"), in viewing a film, actively practice materials covered in the audio-visual presentation during prescribed intervals between sections of the film. When students are tested after being subjected to such an active-review procedure, immediate retention has been shown to be higher than when an equal amount of film time is devoted to passive review (Hovland, Lumsdaine, and Sheffield, 1949; Lumsdaine, May, and Hadsell, 1947, 1958.) Two possible explanations of this phenomenon were offered by the previous investigators. The first is that the active-review procedure guarantees increased practice of the correct responses, since in the passive-review procedures there is no assurance that subjects really practice. The second hypothesis offered was that the active-review procedure increases overall motivation or set to learn and results, therefore, in improved acquisition and performance.

The present study was designed to provide data on the relative contributions to learning of (1) motivation to learn arising from the participation situation itself, and (2) added and emphasized practice during participation sessions. If practice *per se* is the only important factor in learning in the participation situation, then there should be no improvement in the learning of items not practiced during the participation sessions beyond what would occur by simply seeing the film through with no chance for "practice-participation." On the other hand, if motivation

[1]The experimental work reported in this chapter was conducted in 1950-51 under Air Force Contract AF 33(038)-22944 with Boston University, monitored by A. A. Lumsdaine. F. D. Sheffield served as consultant during the initial planning of the research. The study was reported in detail in an HRRL memorandum report—Michael (and Maccoby), 1951—and a somewhat condensed account of the study was subsequently published by Michael and Maccoby (1953). The present chapter was prepared by Maccoby as a revision of the 1951 report.

arising from the participation sessions is effective as a device for increasing learning, then those items not practiced should also improve, since presumably such motivation would constitute a generalized participation effect, having consequences for all pertinent stimuli to which the subject is exposed [e.g., through increasing his level of attentiveness].

Two other sets of variables were also investigated: (a) provision of knowledge of the correct response (KCR) after practice versus no such knowledge (no-KCR), and (b) "overt" versus "covert" practice. Further, it seemed reasonable to assume that the motivational effects of an active review situation would be more likely to be observable if other extrinsic sources of motivation were relatively low. Therefore, an attempt was made to investigate the interactions of these variables at two levels of extrinsic motivation. It was predicted on the basis of previous research results (Hovland, Lumsdaine, and Sheffield, 1949) that the expectation of a test would produce a higher level of extrinsic motivation than would exist without this expectation. By studying the interactions of these variables, in addition to their separate contributions, considerable information could be obtained on the dynamics of the active-participation process.

On the basis of previous related research and theory (Arps, 1920; Carr, 1930; Hovland, 1951; Judd, 1905) the provision of the correct response following participation was expected to enhance the learning benefits of participation, primarily by providing opportunity for the correction of wrong answers and/or further opportunity for covert rehearsal of the correct response. The participation situation was varied by providing practice/participation followed by KCR, and participation without KCR. These conditions provided additional information on the comparative contributions of motivation arising from participation and that arising from practice in the presence of KCR.

In previous experiments, audience-participation procedures fostered overt responding (e.g., Hovland, Lumsdaine, and Sheffield, 1949; Lumsdaine, May, and Hadsell, 1947, 1958). There appear to have been no pertinent experiments on the relative contributions to the efficiency of learning of written practice of meaningful material compared to practice by simply "thinking" the material. If covert participation is equally as effective as overt participation, it would often be much more practicable to employ covert participation in training-film showings.

Method

The Film

The film used was an edited color version of Pattern for Survival (distributed by Cornell Pictures), running 13 minutes, 44 seconds. The subject matter was civilian defense against atomic bombing. (See Supplementary Table 1 at end of chapter for editing schedule.)

Principal Experimental Variations

The design provided that the three experimentally manipulated factors, overt versus covert, KCR versus no-KCR, and announcement of a test versus no announcement of a test, would be arranged so that the eight participation groups would represent all eight possible combinations of the three factors.

Motivation Versus Practice Effects. The relative contributions to learning of practice and motivation under participation conditions were investigated in the following way. During each participation session, half of the 30 final test items covering the previous section of film were specifically practiced, and half of them were not. This procedure was followed for all combinations of the other factors for all experimental treatments. The relative contributions of practice and motivation for each experimental class could then be determined by comparing the test scores on the 15 practiced items versus the 15 non-practiced items, and by comparing the scores on the same items obtained by control groups who saw the film but did not have participation sessions.

Twelve groups of subjects were shown a film, eight of the groups under conditions of audience participation. Each of the twelve groups contained approximately four classes of about 20 high school juniors or seniors. In some treatments the N's obtained for four classes were considered too small, so additional classes were obtained. In all, a total of 1,029 subjects was used. Approximately two minutes after the end of the film, they were tested on their knowledge of the factual material covered by the film. The film and test were administered to subjects in the rooms in their high schools usually used for visual-aids purposes. The assignment of classes to different experimental treatments was done on a random basis.

Overt and Covert Participation Procedures. In each of the eight experimental groups there were four participation sessions. These occurred during three breaks in the film and at the end of the film. Participation took the form of subjects answering orally administered practice questions covering some of the points presented in the preceding section of the film. In the overt-practice groups, subjects wrote their answers to the practice questions on answer sheets. In the covert groups, subjects were instructed just to "think" the answers. Instructions to all groups stressed that the participation periods were to help them learn the film content (i.e., not to test them). It was emphasized that the participation material covered only part of the film content.

Knowledge of the Correct Response. Half of the experimental classes were provided with KCR after answering the participation questions, and half were not. The subjects were informed of the correct answer immediately after they had had an opportunity to answer each question.

In the eight participation groups a total of 11 seconds of participation was provided for each question. For those groups not informed of the correct response, the entire 11 seconds were available for uninterrupted practice, i.e., for attempting to answer the question. In the KCR groups nine seconds were allotted to initial practice, and the remaining two seconds were available for rehearsal of the correct response.

Other Features of Design and Procedure

Test. It was announced in advance to half of the groups that there would be a test at the end of the film; no announcement of a test was given to the other half of the groups. However, the groups that were told in advance that they would be tested were not told what points the test would cover. The test contained 30 orally administered questions divided into two sub-sets, designated as Form 1 and Form 2, selected to be as nearly equivalent as possible with respect to difficulty of the items and distribution in the film. For half of the classes in each experimental treatment the Form 1 items were the ones practiced, with the Form 2 items comprising the non-practiced ones. For the other half of the classes in each group the reverse was true. The combination of the test-announcement factor with the other three variables discussed above involved the design shown in Table 1.

TABLE 1

Basic Experimental Design

	Participation Conditions			
	Overt Participation		Covert Participation	
	KCR	No KCR	KCR	No KCR
Test announced in advance	5 classes*	5 classes*	4 classes	4 classes
Test NOT announced in advance	4 classes	4 classes	4 classes	4 classes

*See Footnote 2, page 276.

Control Groups. In addition to the eight experimental groups, four control groups were employed. First Control Group (Group 1 in Table 2): In order to determine the initial level of information on the material covered by the film, four classes were tested prior to the showing of the film. This provided a base-line against which to compare the teaching efficiency of the film and of the various participation methods. Second Control Group (Group 2 in Table 2): This group of four classes was shown the film without any participation and was tested afterwards. Several fifteen-second pauses were allowed between sections of films by inserting 15 seconds of blank leader in the films at each of the points where they were interrupted for the participation groups. Third Control Group (Group 3 in Table 2): This group of four classes was the same as Group 2 except that subjects were told in advance that they would be tested at the end of the film on its contents.

Control Groups 2 and 3 provided a base-line so that the effectiveness of any of the audience-participation procedures used during the presentation of the film could be compared with that of viewing the film without the use of any participation procedure.

Fourth Control Group (Group 12 in Table 2): It was possible that the non-practiced items might be meaningfully related to the practiced items. If so, there might be some information in the answers to the practiced items that could be applied to answering the non-practiced items. Thus, some or all of the gains in the non-practiced items in the student-participation groups, as compared to the non-participation groups, might be due to transfer effects of practice rather than to increased motivation to learn. A careful comparison of content did not reveal any plausible meaningful relations between one set of items and the other. Nevertheless, this fourth control group (Group 12) was run as a check on the possibility of gains due to transfer from practiced to non-practiced items. In this group all practiced questions were asked at the end of the film rather than at intervals during the course of the film. This ruled out the possibility of the participation procedure's motivating the students to learn more during the showing, since the film had already been seen before the participation/practice took place. If there was to be transference of meaningfully related material, there should have been an increase in test performance on the unpracticed material in Group-12 classes (compared to Groups 2 and 3), even with no participation-derived motivation to encourage this increased learning. However, if no gains on the non-practiced items were to occur in this group, the plausibility of contamination due to transference from the practiced items would be greatly reduced.

Group Sampling. Experimental treatments not only were administered to class units rather than individuals, but also were assigned to pre-formed class groups, taking the classes as they came. Consequently, for purposes of estimating experimental error in group-to-group comparisons, the sampling unit was the class, not the individual. For these comparisons, class means rather than individual scores were the basic units of observation for sampling, and degrees of freedom were obtained from the number of classes rather than number of individuals. However, this group-sampling restriction does not logically apply to those analyses that compare practiced versus non-practiced items within experimental groups. Here, selective factors due to the fact of group sampling would not bias the comparison (except possibly in terms of unlikely-seeming interaction effects), and N could thus be based on number of subjects rather than number of classes.

It should be noted that four separate replications of each of the main eleven groups were run successively. Within each replication the classes were assigned to the eleven treatments on a random basis. Experimenters were rotated in order to equalize possible differential instructor-experimenter factors.

Scoring and Computation of Scores. The test consisted of items of information directly covered in the film. All questions were of the open-end variety which could be answered with very few words. A possible total of six points was allowed on each item (with part scores of two, three, or four points possible on many). The maximum possible score on the whole 30-item test was thus 180; the maximum possible score was 90 on the 15 practiced items, and 90 on the 15 non-practiced items. High reliability of scoring was insured by a preliminary procedure involving the independent practice-scoring of the same sample of papers by each of the several scorers. Differences were located and discussed. This process was then repeated on a new sample until consistent scoring standards were attained. After final scoring began, a random sample of one in ten papers was independently re-scored. The ratio of agreements to total possible agreements was computed in this way for each item. These ratios ranged from 91 per cent to 100 per cent, with a mean of 96 per cent for the whole test.

Control Groups 1, 2, and 3 were also tested on the same 30 items (without having had practice on any of them). In order to compare these groups with the experimental groups prior to calculation of final percentage scores, it was convenient to convert the scores of Control Groups 1, 2, and 3 to a base of 90 points. (This was necessary because all members of the other experimental groups practiced on 15 items, leaving the remaining 15 as non-practiced items.) Therefore, the mean scores for classes in Groups 1, 2, and 3 were determined by separating their total scores (composed of Form 1 and Form 2 items) into sub-scores for Forms 1 and 2, finding the mean of each form score, and then taking the mean of the two means.

Test Administration. In all, 49 classes were run, 48 of which are included in the results.[2] These comprised a total of 1,029 students. IQ information was obtained on 922 of these. In all eight of the experimental groups, as well as in Groups 2 and 3, the final test was administered immediately after the film was shown. In Group 12 the test immediately followed the post-film participation session, while in Group 1 the test preceded the showing of the film.

As already noted, this final test consisted of 15 items that had been practiced during the four participation sessions, and 15 items that had not been previously practiced but were covered by various portions of the film. The total set of 30 items was divided into two sub-sets designated as Form 1 and Form 2, selected to be as nearly equivalent as possible with respect to difficulty of the items, distribution of locations where they were covered in the film, etc. For half of the classes in each participation group, the Form 1 items were the ones practiced, with the Form 2 items comprising the non-practiced ones. For the other half of the classes in each group the reverse was true (Form 2 used as practiced items, Form 1 as non-practiced items). This counter-balancing arrangement was designed to cancel out possible inequalities in inherent difficulty of the practiced and non-practiced items, and thus to permit a more clear-cut comparison to assess the relative contributions of practice versus "motivation" or "set" factors to the effects of audience participation.

[2]Because of the small N's obtained in Groups 5, 7, and 10 after running the first 44 classes, an additional class was run in each of these groups. In Group 10, one class was later dropped because the procedural records were incomplete for this class. Thus four classes remain in Group 10.

Under the conditions described, if the performance of the participation groups turned out to be superior to that of the non-participation groups, only on the items actually practiced during the participation sessions, it was considered that the audience participation effects should be attributed to the direct effects of practice provided by the participation. If, however, there was also an improvement in the non-practiced items due to the use of participation procedures, such improvement would—particularly if not obtained in Group 12—be ascribed to the operation of "motivational" or "set" factors. Finally, in the latter event, the operation of straight practice effect, above and beyond motivational or set factors, would be demonstrated only if the improvement on practiced items was reliably greater than that on non-practiced items.[3]

Resumé of Experimental Design

A more complete schematization of the design is suggested by the layout on the standard tabulation sheet used for summarizing results during analysis of the experimental data, presented as Table 2. In this layout, note that a "cell" of the design is provided not only for the 12 experimental groups, but also for the sub-scores on practiced versus non-practiced items within each group. For purposes of investigating the motivational versus practice bases for effects of participation, comparisons between those intra-group cells are logically as much a part of the design as are group-to-group comparisons for investigating other factors such as overtness of response. The total design may thus be thought of as involving 21 "cells." Means for three of these (Groups 1, 2, and 3) are each based on responses to 30 test items by four classes per cell; means for 16 cells (practiced items for Groups 4 through 11, and non-practiced items for Groups 4 through 11) are based on responses to 15 items each for two classes per cell.

[3]In computing mean scores for the non-practiced items, so-called "reduced" scores were also calculated. These scores, unlike those given in the rest of this report, are based only on the non-practiced items covered in the last three sections of the film. This was done because, since items covering the first section of the film pertain to information presented in the film before the first participation period, any motivation effects arising from the participation procedure would have had no chance of influencing the learning of material presented in the first section of the film. Since potential effects on non-practiced items due to motivational factors are of importance in this experiment, it appeared advisable to check the possibility that increased sensitivity might be obtained by eliminating that part of the variance in the score introduced by items in the first section of the film. However, in all instances the results of analyses using these "reduced" scores led to the same results as those presented in the report.

TABLE 2

The Twelve Groups Used in the Complete Experimental Design
(Showing Number of Test Items Entering into Score for Each Cell)

Experimental conditions	Control Groups 1, 2, 3 (No Participation)			Experimental (Participation) Groups								Control Group 4 Overt, Post-Film-Participation
				Overt				Covert				
	No film	Film		No KCR		KCR		No KCR		KCR		KCR
Prior Announcement of Test:*		No Test*	Test	No Test*	Test	No Test*	Test	No Test*	Test	No Test*	Test	Test
Group No.	1	2	3	4	5	6	7	8	9	10	11	12
Practiced items	none	none	none	15	15	15	15	15	15	15	15	15
Non-practiced items	30	30	30	15	15	15	15	15	15	15	15	15
Totals	30	30	30	30	30	30	30	30	30	30	30	30

*"No Test" means test was not announced in advance.

Results[4]

General Effects of the Film

Initial Ability. The mean test score of the four classes in Group 1, tested before being shown the film, was 22.9 correct answers (or 25.4 per cent). Correct answers on each question, however, were given by some individuals; in other words, there was no single item for which the mean score was zero before the film.

Gains Produced by the Film Alone. Control Groups 2 and 3 were shown the edited film, but with no provision for audience participation. These groups, of four classes each, obtained mean test scores which were 50.9 per cent and 54.1 per cent respectively of the total possible test score. When the population is divided into high and low IQ sub-groups[5] (the dividing line being the mean population IQ), the differences in absolute percentage points between the no-film and film groups is somewhat greater for the higher IQ category, as would be expected. (See Table 3). Although the film as such did not appear to differ markedly in its effects on

TABLE 3

Mean Raw Scores for Control Groups 1, 2, and 3
(Mean of 30 Items Divided by 2)*

	No Film	Film Only		Av. of "Film Only" Groups
	Group 1	Group 2	Group 3	Groups 2 & 3
All persons	22.9	45.8	48.7	47.3
High IQ	23.5	53.5	50.4	52.0
Low IQ	18.2	39.7	45.7	42.7

*Mean for 30 items is divided by two to reduce to a base of 90 points for comparability with subsequent data for 15-item scores, as explained on page 276. The basic data on which this and other comparisons for high- and low- IQ sub-groups are derived are presented in Supplementary Table 2.

[4]For basic data tabulations underlying the results, see Supplementary Table 2 at the end of the chapter (cf. also the summary table given by Michael and Maccoby, 1953, p. 415). Unless otherwise indicated, all values given are based on means of the class-means.

[5]In the case of IQ breakdowns, class-means were computed separately for each of the two IQ sub-groups. A few cases for which no IQ data were available are included in the means for "all persons."

the high and low IQ components of the population, IQ does enter as a factor influencing the level of test performance. When individual test scores and IQ for a given experimental condition are correlated, the results average about $r = .5$. Separate analyses for high and low IQ groups are accordingly included in the analysis of the influence of the several experimental factors.

Range of Performance. The lowest mean score for any of the groups is that of Group 1, which is 25.4 per cent of the possible score. The highest mean score attained was by the "post-film-practice" group (Group 12) on the practiced items—82.8 per cent of the possible score. The non-participation film classes in Group 2 had a mean score of 50.9 per cent. Thus there appear to be suitable conditions for efficient experimental measurement—i.e., minimum conditions for the occurrence of either ceiling or floor effects or for clustering of score averages due to too limited a range. (See Supplementary Table 2).

Effects of Student-Participation Procedures

Overall Gains Produced by Student Participation. The participation procedures used in this study produced statistically significant gains in learning, over that produced by simply viewing the film. Comparison of mean percentage scores (mean of class means) based on all items, for control groups 2 and 3 (film with no participation) versus experimental conditions 4 through 11 (participation) shows a mean per cent of 59.8 per cent for groups with participation, versus a 52.5 per cent for groups with no participation. This gain of 7.3 per cent due to the participation procedure is reliable at between .01 and .02 probability levels.[6] In terms of this overall gain, the effects of the participation procedure used here are similar to those found in previous research described (cf. Hovland, Lumsdaine, and Sheffield, 1949, Ch. 9). Table 4-A shows the same comparison for high and low IQ sub-groups separately.

TABLE 4-A

Average Effect of Participation Procedures for High and Low IQ Subgroups (In Terms of Mean Percent of Correct Answers on all Test Items)

	Participation (Combined Groups 4-11)	No Participation (Combined Groups 2 and 3)	Difference	t	P
High IQ	64.3%	57.7%	6.6%	2.5	<.02
Low IQ	55.7%	47.4%	8.3%	3.2	<.01

[6]All P-values given in this report are for both tails.

TABLE 4-B

Effects of Participation Procedure
on Practiced and Non-Practiced Items for
High and Low IQ Subgroups
(Mean Percent of Correct Answers)

	Participation Groups	Non-Participation* Groups	Difference	t	P
High IQ					
Items Practiced by Participation Groups	73.1%	57.7%	15.4%	3.7	<.01
Non-Practiced Items	55.5%	57.7%	-2.2%	-.8	-
Low IQ					
Items Practiced by Participation Groups	64.3%	47.4%	16.9%	4.4	<.01
Non-Practiced Items	47.1%	47.4%	-.3%	.1	-

*As previously explained, the no-participation scores are means based on all 30 items, since the absence of participation in these classes provided no opportunity for practice on any of the items. The means of the participation groups are each based on 15 items, since in all instances 15 items were practiced and 15 were not. The particular items selected for practice were always counterbalanced—that is, half of the classes in any given experimental condition practiced on Form 1 and half practiced on Form 2.

Since the higher base-line (no-participation level) for the high-IQ sub-groups allows somewhat less room for improvement, it is also of interest to express these results in terms of gain relative to room for possible improvement over the level reached by the film alone. Computing an "effectiveness index" based on the average values in Table 4-A yields values of 15.5 per cent for the high IQ group and 15.8 per cent for the low IQ group.[7]

[7] Effectiveness Index = $\frac{\text{test score (\%) - control score (\%)}}{\text{100 - control score (\%)}}$.

This index expresses gains as a ratio relative to amount of room for improvement—i.e., ratio of actual gain to maximum possible gain. For further discussion see Appendix A in Hovland, Lumsdaine, and Sheffield (1949).

Practice Versus Non-Practice of Items to be Learned. The comparison of gains in this experiment with the results of the previous experiment is more nearly parallel when data are compared for only the items actually practiced, since in the previous research all items were practiced (Hovland, Lumsdaine, and Sheffield, 1949). The superior results of participation procedures were found only in connection with those items that were actually practiced during the participation sessions; the items not specifically practiced did not appear to be learned any better than they would have been by viewing the film alone. (See Figure 1.) The mean number of correct answers, for Groups 4 - 11 combined, is 51.3 per cent for those items that were not practiced during the participation sessions, as against 68.3 per cent for the items that were practiced during the participation sessions. As shown in Figure 1, the performance of participation groups on non-practiced items was substantially the same as the average performance of those who saw the film without any participation sessions. A similar comparison is made for high and low IQ groups separately in Table 4-B.

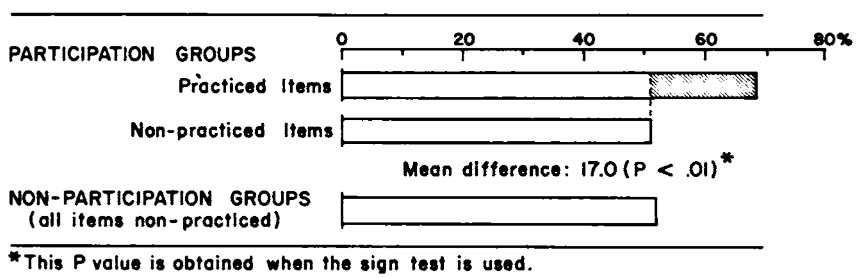


Figure 1. Average Per Cent of Correct Answers for Participation Groups on Practiced Items and on Non-Practiced Items, Compared with Average Performance of Non-Participation Groups.

Influence of Various Experimental Factors on the Effectiveness of Participation Procedures[8]

Providing Knowledge of the Correct Response Following Practice (KCR versus No-KCR[9]). As shown by means for practiced items, there was an average gain of

---

[8]See Supplementary Table 3 at end of chapter for summary of mean scores entering into the following analyses.

[9]The term "feedback" (cf. Wiener, 1948) and other related expressions such as "knowledge of results" are in a sense misnomers, as applied to procedures of the kind used here, in that the information provided following practice is the same, from the stimulus point of view, regardless of what response the subject has made during practice. In the knowledge-of-results procedure as traditionally applied to practice at motor skills, by contrast, the information fed back depends on the extent and kind of errors just made in practice. In both instances, however, the "feedback" does provide a basis either for recognizing successful (correct) performance or for recognizing and correcting errors made during practice.

16.3 percentage points from the condition of No-KCR (mean score, 76.5 per cent for groups 6, 7, 10 and 11) to the condition of KCR (60.2 per cent). This is a clearly significant difference (P <.01); the provision of the knowledge-of-the-correct-response procedure thus appears to be a sizable contributor to learning over and above the existence of the basic kind of participation utilized in this experiment. The value of t for the 16.3 per cent difference between KCR and No-KCR was 6.4 (P <.01). The gains due to the KCR procedure are also clearly in evidence when the data are broken into high and low IQ sub-groups. The following tabulation, derived from basic data summarized in Supplementary Table 2, shows the statistical comparisons, for the two IQ sub-groups, of differences between the KCR and no-KCR conditions:

	Difference	t	P
High IQ	17.4%	7.0	<.01
Low IQ	16.4%	7.5	<.01

Effectiveness of the No-KCR Condition of Participation. In spite of the marked superiority of participation with KCR over the procedure without KCR, comparison of the overall value of 60.2 per cent for the latter, as against only 52.5 per cent for the film-only groups, suggests that even without KCR the participation procedure can be effective in improving performance on the practiced items. The value of t for this difference of 7.7 per cent is 2.6 (P <.02). However, when high and low IQ groups are analyzed separately, the effects of participation without KCR are found to be significantly superior at a high level of confidence only for the low IQ classes:

Difference Between Scores of No-KCR
Participation Groups vs. No-Participation Groups

High IQ			Low IQ		
Diff.	t	P	Diff.	t	P
6.4%	2.0	<.06	8.6%	3.3	<.01

Overt (Written) Versus Covert ("Mental") Practice. The effect of the overt-covert variable on learning when the other variables are lumped together is seen in the mean percentage of correct answers for practiced items of 70.4 per cent for the overt groups (4 - 7) versus that of 66.1 per cent for the covert groups (8 - 11). Using the class as the sampling unit, the difference of 4.3% between the amount learned by overt practice and that learned by covert practice is not significant (.3 > P >.2). When the subjects are divided on the basis of IQ, there is also no significant difference between overt and covert learning scores, for either IQ subgroup, in terms of class-sampling tests.[10] Comparisons of the gains due to overt participation and those due to covert participation, as against no participation, are summarized in Table 5.

Joint Effects of Overt or Covert Participation and KCR or No-KCR. It is evident from Table 6 that KCR is effective regardless of whether it is used with overt or covert participation procedures, for both high and low IQ sub-groups.

TABLE 5*

Effects of Overt and of Covert Participation
Overall and for High and Low IQ Sub-Groups
(In Terms of Mean Per Cent of Correct Answers on Practiced Items)

Means	No Participation	Overt Participation	Covert Participation
Overall	52.5%	70.4	66.1
High IQ	57.7%	74.0	72.0
Low IQ	47.4%	65.3	63.1

Differences	Overt vs. No Participation			Covert vs. No Participation		
	Diff.	t	P	Diff.	t	P
Overall	17.9	4.4	<.01	13.6	3.2	<.01
High IQ	16.3	4.2	<.01	14.3	3.0	<.01
Low IQ	17.9	4.7	<.01	15.7	3.7	<.01

*Note that this table includes both KCR and No-KCR contributions to Overt and Covert participation. (cf. Table 6).

[10]When the groups are split on the basis of IQ, if the individual were used as the sampling unit, the low IQ people would appear to do significantly better when they practice overtly (.05 > P > .03) than when they practice covertly. However, when the low IQ scores are corrected for IQ differences by a covariance adjustment, the P value becomes approximately .06. A similar calculation for the difference in performance for the high IQ people for the two conditions would lead to a less "significant" P value (between .2 and .3). See, however, the earlier remarks on group sampling.

TABLE 6

Separate Comparisons of KCR Versus No-KCR Groups Under Overt and Under Covert Participation Conditions

	All Persons	High IQ	Low IQ
Overt			
KCR	78.7%	82.3%	73.9%
No-KCR	62.2%	65.9%	56.7%
Difference	16.5%	16.4%	17.2%
Covert			
KCR	74.1%	80.7%	70.8%
No-KCR	58.1%	62.0%	55.4%
Difference	16.0%	18.7%	15.4%

The fact that under KCR conditions the covert participation procedure appears to be about as effective as the overt procedure has important implications that will be discussed below.

It was shown above that participation without KCR was less effective than with KCR, for both high and low IQ sub-groups. Also, it was pointed out that when all conditions of no-KCR are lumped, there is a significant increase in learning compared to the no-participation groups. Table 7 further indicates, however, that a significant improvement under covert participation conditions in the no-KCR groups was obtained only in the low IQ groups.

TABLE 7

Increase in Scores of Overt Participation (with no KCR) and Covert Participation (with no KCR) Compared to No-Participation Groups

	I.Q.	Diff.*	t	P
Overt, No-KCR	High	8.1%	2.6	<.05
	Low	9.2%	3.8	<.01
Covert, No-KCR	High	4.2%	1.0	<.4
	Low	8.0%	2.2	<.05

*Difference between the per cent of correct answers for the indicated No-KCR participation group and the per cent correct with no participation procedure.

Analysis of Other Experimental Variations

Test-Announced vs. No-Test-Announced Before the Film. A further comparison is that between groups that were and those that were not informed in advance that they would be tested on the content of the film. When the other variables are lumped, the difference between the mean percentages of correct answers for all items given under the test-announced condition (59.9 per cent) and under the test-unannounced condition (59.7 per cent) is not significant. Essentially the same results obtain when the non-practiced and practiced items are considered separately, and when the groups are sub-divided in terms of high and low IQ for practiced items. (Approximate percentages for practiced items with versus without test announcement, respectively, are 77 per cent versus 73 per cent for high IQ, and 67 per cent versus 64 per cent for low IQ.)

It had originally been hoped that announcement of the test would provide some differentiation in level of extrinsic motivation (reflected in higher scores on both practiced and non-practiced items, as had been earlier obtained by Hovland, Lumsdaine, and Sheffield). In this case, analysis of data might confirm the expectation that the lower motivation group (test not announced) having the greatest opportunity to do so, would accordingly, show gains on the non-practiced items, interpretable as evidence of motivational effects due to the participation procedure. This hope was frustrated, however, since the non-practiced items were not significantly influenced overall by the participation procedure, and also because no significant overall differences occurred between the "test-announced" and "test-not-announced" groups. [See Chapter 19 for further investigation of this point.]

Results for Group 12. The results for Group 12 were obtained, it will be recalled, to be used as a check on the possibility of "cognitive" (as distinct from "motivational") carry-over in the event that non-practiced items showed gains from the participation procedure. Since this was not the case, this control turned out to be unnecessary for its original purpose. However, it is of interest to examine the results of Group 12 because of the fact that this procedure apparently produced highly effective learning on the practiced items. The overall values for Group 12, as compared with the corresponding data for Group 7, are shown below. The comparison is made with Group 7 because Group 7 was the nearest counterpart to Group 12 among the regular participation procedures (overt participation with KCR and with announcement of a test in advance).

Comparison of Group 12 with Group 7
(mean per cent correct for practiced items only)

Group 12	82.8%
Group 7	78.6%

Discussion

The results obtained in this experiment extend the findings of earlier research with respect to the efficacy of practice provided by audience participation in learning simple and complex verbal material. There is good reason, then, to propose that, at least in situations where verbal material is to be learned, student participation can be used as a means of increasing the amount of learning over and above that which would be learned without such participation.

Since better test performance by the experimental groups (as compared to the control groups) was not found on those items which were not practiced during the participation sessions, it is concluded that the use of the participation procedures did not produce the kind of motivation which results in increased learning of film material not specifically practiced.

At the present time the best generalization from available evidence, including both the data of the present experiment and that of other related experiments, would appear to be as follows: When participation procedures of the sort described in this research are employed with "discrete" responses, such as items of factual information under generally high conditions of motivation, the improved learning derives primarily from the conditions of practice during participation and not from motivation to learn generated by the participation situation. The effects are confined to the material specifically practiced. However, where the material practiced in the participation sessions consists of examples of a principle or method which is to be learned, carry-over to similar, non-practiced material can be expected (e.g., cf. the pertinent data presented by Kimble and Wulff in Chapter 15). Whether such motivational carry-over occurs on discrete factual material under conditions of generally low motivation was not tested in this study (see, however, further data on this point reported in Chapter 19). It is reasonable to assume that the subject matter and mode of presentation in this film generated a high level of motivation to learn for most of the subjects. If this was the case, then additional incentives to learn, of the sort provided in this experiment (e.g., the test announcement), may have been incapable of adding a significant motivational increment.

The most important contributor to effective participation, under the conditions of participation studied, is KCR. When no such knowledge is available, the improvement in learning is considerably reduced.

Here, it seems, is a highly general finding: The proposition that superior learning will occur for either manual or verbal activities when the learner is given the opportunity to rectify an error by knowing that an error has been made, what the error was, and what the correct behavior should be, has not only the empirical support of this experiment, but considerable support from many studies of learning and previous experimentation in other media, as well as in films (Hovland, 1951).

However, improvements in learning, while not nearly as great as with KCR, occur without KCR. These improvements can be explained by the operation of a recency effect, as follows: Those subjects who practiced the <u>right</u> answer during the practice periods are more likely to know the right answer on the final test than if they had not practiced it, or if no practice period was available. There is greater likelihood that the right answer will be remembered by <u>more</u> people at the time of the practice period than would remember it if not exposed to the question again until the test period. Therefore, even the practice participation periods which do not have KCR are of value because they can at least take advantage of recency.[11]

In the present experiment, overt practice did not lead to appreciably better learning than did covert practice. Further, the combination of either covert or overt practice with KCR accounted for substantial, and approximately equal, increases in learning and proved to yield the most effective participation conditions employed in this study.

While there is necessarily no supporting evidence available from this study in the covert practice case, there is no reason to assume that, up to the time of KCR, the percentage of subjects knowing the right answer is different among those who practice covertly from the percentage among those who practice overtly. Since there is no significant difference between "overt no-KCR" and "covert no-KCR," the methods appear equally efficacious in consolidating the correct response previous to KCR.

Whatever increments of learning occur after KCR would be the result of the same kind of practice that occurred in both the overt and covert groups; namely, two seconds of covert practice before the next question is asked. Therefore, the first nine seconds of the practice trial could be expected to influence learning during the last two seconds of practice only if there were some interaction between the practice during the last two seconds after KCR and the kind of practice, overt or covert, occurring during the first nine seconds before KCR. The results indicate no such interaction, and the final scores for the two forms of practice are not significantly different. Thus, it can probably be concluded that the chief importance of the practice period, overt or covert, before the KCR is two-fold: (a) to permit practice that will help fix the correct answer for those already able to give the correct response at that time; and (b) to get the student sufficiently motivated to learn the response, so that when the KCR is given, he will practice it during the succeeding two seconds of covert participation.

The comparative results for overt and covert practice seem especially important from the standpoint of the design and utilization of instructional media. When overt practice procedures are employed, additional practice materials and special conditions and assistance for proper utilization of these materials are necessary. It would appear, therefore, that all the conditions for substantially increasing learning could presumably be incorporated directly into a film at fairly

[11]More elaborate inter-score and intra-score comparisons than those presented in this paper indicate that most of the improvement in learning for the no-KCR groups can logically be accounted for by the operation of the recency effect. However, small but evidently real learning increases occur on items practiced in Practice Periods 3 and 4 that cannot logically be accounted for by the operation of the recency effect. These increases have been ascribed to artifacts in the design, and not to the participation sessions as such.

little cost, and no special classroom procedure or materials would be necessary. Further experimentation is necessary, however, to confirm the expectation that provisions for covert practice built into the film itself would be effective as compared to the "live" conditions utilized in this experiment.

Summary

Problem

Previous research has shown that more learning from films occurs when the viewer is required to take part in an "audience participation" procedure than when he is not. This increased learning might be the result of (a) increased practice during the participation periods, and/or (b) increased motivation to learn arising from participation. The chief purpose of this experiment was to assess the relative contributions of practice and motivation factors to the increased teaching efficiency of film showings using a participation procedure. In addition, three experimental conditions were studied that were expected to affect the level of learning.

Participation Conditions

Participation occurred during practice periods alternated with sections of the film, and required the audience to answer questions asked on the content of the previous section of the film. Subjects participated by trying to answer the questions, some by writing the answers (overt practice) and some by just "thinking" them (covert practice). Under each of these conditions, half of the classes were told the correct answers after they had tried to answer the questions themselves (KCR), and half were not. Half of the classes under each pair of these conditions were told before the film began that they would be tested afterward; the other classes were not. Control groups consisted of subjects who were given the test questions but not the film, and subjects who saw the film straight through without participation (half of whom were warned they would be tested on the film content, and half of whom were not warned).

Findings

1. Within the limitations implicit in this experiment, audience participation procedures utilizing either overt or covert practice, when the participants had knowledge of the correct responses (KCR), were found to result in considerable improvement in the learning of verbal material over that resulting from simply viewing the film. Even without the provision of the correct response, participation appears to result in better performance than no participation. This increase in learning seems to be due primarily to the effects of practice and not to the effects of changes in motivation to learn.

2. The most important factor influencing the amount of learning in this experiment was the provision of knowledge of the correct response after practice (KCR). This applies to both overt- and covert-practice conditions, and to both the more intelligent and less intelligent participants.

3. There was no significant difference between the level of learning achieved with overt practice as compared to covert practice. An unsuccessful attempt was made to provide two levels of extrinsic motivation by informing some subjects of a test to follow the film, and not informing others; no significant differences were found.

SUPPLEMENTARY TABLE 1

Editing Schedule for Experimental Versions of "Pattern for Survival"*

	Content	Cumulative running time
1.	8 sec. introductory music to "pling"	0'8"
2.	Delete 8 sec.+ to 4'38" including "Thank you, Bill Laurence."	---
3.	1st section: 2'48" from 4'38" to and including "imbedded in the ground."	2'56"
4.	2nd section: 1'30" beginning "now that we know. . .", ending "if they react quickly."	4'26"
5.	3rd section: 3'04" beginning "at this point you're going to ask", ending "spare batteries."	7'30"
6.	4th section: 6'12" beginning "those are the material preparations", ending "help those who cannot help themselves."	13'42"
7.	Delete from "their's will be a noble duty" for 55".	---
8.	Close with 2" of closing music including "The End" title.	13'44"

*Specified in minutes (') and seconds (") from designated reference points.

SUPPLEMENTARY TABLE 2-A

Basic Data Table Showing Class Means and Overall Means for Each Treatment
(Mean of Class Means for Raw Scores: Maximum Possible Score = 90)

	Class Number	Form 1** For Groups 1-3 (See Table 2-B for Form 2 Scores)			Table 2-A** Practiced-Item Scores for Groups 4-11*								
		1	No Test* 2	Test 3	No Test* 4	Test 5	No Test* 6	Test 7	No Test* 8	Test 9	No Test* 10	Test 11	Test 12
OVERALL	1	23.8	40.1	52.4	49.7	52.3	77.3	64.7	44.3	56.7	59.4	72.2	73.5
	2	23.5	41.4	44.3	54.3	62.3	64.2	74.1	61.8	54.4	68.3	65.7	-
	3	17.8	47.6	52.1	67.3	51.2	76.2	80.3	42.3	45.9	-	61.2	75.6
	4	26.8	58.6	58.0	56.5	49.5	65.8	68.1	54.3	58.4	57.5	74.7	-
	5					60.8		66.8			74.8		
	Mean	23.0	47.9	51.7	57.0	55.2	70.9	70.8	50.7	53.9	65.0	68.4	74.6
HIGH I.Q.	1	24.9	46.5	53.6	52.6	55.5	79.2	69.7	50.7	69.1	59.0	80.8	74.8
	2	27.2	46.2	48.9	56.1	62.9	73.5	76.5	62.1	54.3	77.3	71.0	-
	3	18.3	59.3	51.3	68.8	55.3	77.2	80.9	46.5	***	-	67.0	79.5
	4	29.0	67.3	61.4	64.3	53.3	68.7	73.3	47.0	60.9	76.0	74.8	-
	5					64.7		67.3			75.6		
	Mean	24.8	54.8	53.8	60.5	58.3	74.6	73.5	51.6	61.4	72.0	73.4	77.2
LOW I.Q.	1	16.8	38.2	49.4	47.5	50.1	69.0	61.0	41.3	48.7	59.6	65.0	72.4
	2	15.6	36.0	38.6	49.6	51.3	62.8	66.5	61.0	54.4	61.7	63.8	-
	3	12.8	43.8	55.2	58.3	49.5	67.5	77.1	40.7	43.9	-	57.3	70.7
	4	22.8	39.0	42.5	50.6	48.4	65.2	63.5	56.6	52.3	54.6	75.2	-
	5					53.7		66.1			73.0		
	Mean	17.0	39.2	46.4	51.5	50.6	66.1	66.3	49.9	49.8	62.2	65.3	71.5

* See first footnote for Table 2-B, on the following page
** See second footnote for Table 2-B
***See third footnote for Table 2-B

SUPPLEMENTARY TABLE 2-B

	Class Number	Form II** For Groups 1-3 (See Table 2-A for Form I Scores)			Non-Practiced-Item Scores for Groups 4-11*								
		1	No Test* 2	Test 3	No Test* 4	Test 5	No Test* 6	Test 7	No Test* 8	Test 9	No Test* 10	Test 11	Test 12
OVERALL	1	26.7	36.3	46.8	48.5	43.9	56.4	43.0	41.8	48.8	46.6	41.8	45.7
	2	20.5	45.3	45.2	47.1	53.0	41.7	43.1	54.6	45.6	46.2	34.6	-
	3	19.4	45.8	44.2	51.4	45.5	53.7	55.8	43.1	38.2	-	39.0	52.2
	4	24.2	51.4	47.0	44.2	39.2	37.6	42.7	43.0	61.5	36.1	50.5	-
	5					56.6		41.0			55.3		
	Means	22.7	44.7	45.8	47.8	47.6	47.4	45.1	45.6	48.5	46.0	41.4	49.0
HIGH I.Q.	1	25.8	44.1	45.2	53.2	52.3	57.6	47.7	46.5	52.1	48.8	43.8	54.0
	2	22.8	51.6	47.1	50.5	52.9	45.0	43.6	57.1	52.2	50.5	43.9	-
	3	19.5	57.7	46.0	51.8	54.8	54.9	53.9	47.8	***	-	42.5	54.5
	4	20.8	55.7	49.6	45.9	49.7	40.7	46.0	40.0	66.5	45.3	56.2	-
	5					58.0		40.5			56.8		
	Means	22.2	52.3	47.0	50.4	53.5	49.5	46.3	47.9	57.0	50.4	46.6	54.2
LOW I.Q.	1	21.3	32.4	50.8	45.3	38.3	57.0	40.6	40.0	44.6	45.8	34.5	38.3
	2	15.2	38.2	42.8	43.0	47.0	39.5	41.5	47.2	38.9	43.0	29.8	-
	3	16.5	47.9	44.7	48.7	39.9	43.0	52.2	41.4	37.5	-	36.7	46.0
	4	24.8	41.8	41.5	42.1	36.0	36.4	39.9	42.6	49.0	35.5	40.0	-
	5					56.4		40.6			52.0		
	Means	19.5	40.1	45.0	44.8	43.5	44.0	43.0	42.8	42.5	43.4	35.2	42.2

* As noted on pages 275 and 276, there were usable data for four classes per group in all groups except Groups 5 and 7 (five classes each) and Group 12 (only two classes).

** As noted in the text, all items were non-practiced for Groups 1, 2 and 3. The entries for these three groups refer merely to the scores for two forms of the test, neither of which was practiced by these three groups.

***Of the four classes in Group 9, one class (no. 3) contained no individuals in the high IQ category.

SUPPLEMENTARY TABLE 3

Mean Test Scores (Per Cent of Correct Answers: Mean of Class-Means)

Groups	Group Numbers	Practiced Items	Non-practiced* Items	Total test score
No film	1	---	25.4	25.4
Straight film (no test)	2	---	50.9	50.9
Straight film (test)	3	---	54.1	54.1
Participation Groups:				
All overt	4-7	70.4	52.1	61.2
All covert	8-11	66.1	50.4	58.2
All KCR	6, 7, 10, 11	76.5	50.0	63.2
All No-KCR	4, 5, 8, 9	60.2	52.7	56.4
All with test announced	5, 7, 9, 11	69.1	50.8	59.9
All with test not announced	4, 6, 8, 10	67.6	51.9	59.7
Post-film Practice Groups**:				
Overt-test-KCR	12	82.8	54.4	68.6
Overt-test-KCR	7	78.6	50.1	64.3
Covert-test-KCR	11	76.0	46.0	61.0

*based on all non-practiced items for each group—i.e., these are <u>not</u> "reduced" scores

**two classes only

CHAPTER 19

FURTHER STUDIES OF STUDENT PARTICIPATION PROCEDURES IN FILM INSTRUCTION: REVIEW AND PREVIEW COVERT PRACTICE, AND MOTIVATIONAL INTERACTIONS[1]

Nathan Maccoby
Donald N. Michael
Seymour Levine

Introduction

This chapter reports two additional experiments on the effects of active-rehearsal, or "student-participation," procedures used in conjunction with film-presented instruction. These consisted of question-and-answer sessions, or similar active-recitation or practice sessions, interspersed between sections of an instructional film.

The first was designed to bear on a number of questions raised by previous experiments, including the study just reported in Chapter 18. In addition to yielding results relevant to several of these questions, Experiment I also replicates in a military situation some findings previously obtained with civilian students, reported in the previous chapter.

Experiment I

Background and Purpose

Pertinent Aspects of Prior Experimentation

It will be recalled from Chapter 18 that: (1) student-participation procedures utilizing either overt or covert practice, when students were then given knowledge

[1]The experimentation reported here, like that in Chapter 18, was conducted at Boston University under Air Force contract AF 33(038)-22944, directed by Nathan Maccoby and monitored by A. A. Lumsdaine (see Footnote 1 of Chapter 18). A report on the first of the two studies here presented was originally circulated in mimeographed form as an unpublished staff research memorandum (Michael and Maccoby, 1954). Contractor's reports on the second study were first prepared by S. Levine in 1953; this study was reported in a paper at the 1953 annual convention of the American Psychological Association and later in a mimeographed Laboratory Note (Levine, 1955). The present chapter was prepared by Maccoby, with the Editor's assistance on some portions, and was based primarily on the two mimeographed papers by Michael and Maccoby and by Levine.

of the correct responses (KCR), resulted in considerable improvement over the learning of verbal material resulting from simply viewing the film; (2) even without the provision of KCR, overt participation appeared to result in superior performance over no participation; (3) significant improvement due to the use of the participation procedures was found only on those points of information rehearsed during the participation sessions, so that the increased learning appeared to be due primarily to the effects of practice and not to changes in motivation to learn; (4) the most important single factor influencing the effectiveness of the participation procedures (either covert or overt) in this experiment appeared to be the provision of knowledge of the correct response after practice; (5) no significant difference was found between the level of learning achieved when overt practice was used as compared to covert practice. An additional, negative result was (6) the absence of a significant difference between the level of learning achieved under the condition of prior warning that a test would follow the film viewing, and the level achieved without this warning. (The negative results (3) and (6) defeated the intention of studying motivational effects of participation as influenced by differences in extrinsic level of motivation.)

Problems Investigated

Consideration of the variables in the study reported in Chapter 18, and the results related to these variables, posed further problems. Those problems chosen for investigation in the present experiment (Experiment I) are listed below. (Specific details of design procedure are discussed subsequently.)

Generality of Findings From Chapter 18. The subjects used in the Chapter 18 study were high school students; those used in this experiment were military personnel. It might seem very unlikely that these two different groups of subjects would be differentially affected by the use of the various student-participation devices; nevertheless, for reasons indicated below, it appeared desirable to check this supposition. Therefore, some of the same experimental conditions that had been run with the high school groups were also run in the new experiment using military groups.

Military Versus Civilian Test Motivation. Previous research has verified the expectation that announcing that a test will follow a learning session can result in an increase in learning over that which occurs when no such announcement is made. In the experiment reported by Hovland, Lumsdaine, and Sheffield, (1949, Chapter 9), prior announcement of a test materially improved the learning scores of military groups. It appeared possible that the apparent lack of effectiveness of this supposed incentive in the high school groups was due to a "ceiling" on motivation for these groups—due to their interest in the subject matter, their habit of expecting tests on class work, etc. If so, with a military audience the motivating effects of announcing a test might show up. In this event, results for the less motivated classes, in which no test was announced, would provide a better check on the expectation that the participation procedure itself could have motivational effects leading to gains on material not directly rehearsed in the participation sessions. In order to check this possibility, half the treatments in this experiment were given with the announcement that a test would follow the film-showing, and half the treatments were given without such an announcement.

Further Replication. It also seemed desirable to compare two experimental conditions that would parallel some of the experimental conditions in the first experiment. The two conditions chosen were covert practice with knowledge of the correct response not given (no KCR) and covert practice with knowledge of the correct response. These two "review-question" procedures were chosen not only because of the large differences between them in the level of learning obtained in the high school groups, but also because they would allow comparisons with a "preview-question" procedure discussed below. Further, it was especially desirable to recheck the efficacy of the covert-participation procedure, because of the practical implications of this method for film design and utilization in instances in which the use of "overt" rehearsal procedures would be unfeasible (cf. later studies of overt versus covert response in teaching-machine programs).

"Rote Learning" Contrasted to "Meaningful Learning." "Meaningful learning" is generally taken to imply that what was learned can be elicited in appropriate situations which are not identical with the one in which the responses were originally learned. Because of the variety of ways in which correct responses were expressed by different individuals on the final test in Chapter 18, it was believed that the learning occurring under participation conditions was probably not describable as purely "rote learning," in spite of the fact that the questions asked during the participation sessions, and their counterparts on the final test, were identical in wording. However, more direct confirmation of the occurrence of "meaningful learning" rather than "rote learning," under participation conditions, was desirable. Evidence for "meaningful learning" of verbal responses would be provided if students could make the proper response to either of two different questions which rationally require the same response. Therefore, in the present experiment some of the questions were worded one way when they were asked during the participation sessions, and another way when they were asked during the final test. (Both sets of questions, however, required the same answers.)

"Preview" Questions Contrasted with "Review" Questions. In Chapter 18, groups that were given the test questions without first seeing the film were then, for the sake of maintaining good relations with the schools, shown the film straight through after having been tested. It seemed to the experimenters that many comments made by the students while watching the film indicated that the pre-film questions tended to encourage greater attention to the film. (Remarks made by the students indicated that they found in the film answers to questions they had been unable to answer during the preceding test.) These observations, and results of previous experiments such as those on effects of attention-directing pre-film tests (see May and Lumsdaine, 1958, Chapter 7), suggested the hypothesis that a participation procedure in which the same questions about a section of film precede each section of film might be effective in increasing learning. Asking the questions before the section of film containing the answers could alert the viewer to specific subject matter and thereby increase the chances of his selecting the pertinent information in the film and learning it better. On the other hand, it appeared possible that with new and unfamiliar subject matter, the trainee might have difficulty in remembering the questions well enough to associate them with the appropriate film content.

Accordingly, it seemed important to investigate the effects of "preview" questions[2] in order to determine how any gains in learning resulting from this method would compare with gains effected by "review" questions following the different sections of film ("review" questions were used in Chapter 18). For purposes of comparison, one experimental condition included presentation of the participation questions before each of the sections of film that contained the answers. Practice was of the "covert" form (simply thinking the answer to each question asked by the instructor) in all experimental conditions and, in the two treatments designed to compare "preview" and "review" participation, the correct answers were not "fed back" directly. It was up to the viewer to select and think about the correct answers from the succeeding (or preceding) section of the film.

Experimental Design And Procedure

General Procedure

Subjects for this experiment included 28 flights of basic indoctrination trainees at Sampson (New York) Air Force Base. The flights were pre-formed; hence the sampling unit was the flight rather than the individual trainee. Flights were assigned at random to the various experimental conditions, with four flights per condition. Assignments of instructors to the different experimental treatments (see below) were rotated to equalize instructor-experimenter factors. The entire experiment was run during a two-day period. Flights in a given experimental condition were run at different times of the day in order to equalize possible differences in overall performance level that might be associated with the time of day.

The film used was Pattern for Survival, edited as indicated in Chapter 18, with the same three film "breaks" for the participation sessions (see Supplementary Table 1, Chapter 18). The film was shown in training school buildings, in rooms customarily used for classroom instruction and training film showings.

The Participation procedure (without "feedback" of the correct answer) consisted of stopping the film, asking a question, allowing 11 seconds for trainees to

[2]The use of "preview" questions designed to motivate students' interest in succeeding sections of a film has been previously studied in an experiment conducted at Yale University (Lumsdaine, May, and Hadsell, 1947, 1958). In the Yale study only one or two "preview" questions were given prior to each section of the film, and these were of a general, diffuse, "curiosity-arousing" nature rather than referring to specific items of information. They differed in content both from the review-participation questions used in the same study and from the test questions used in the post-film test. Evidence was found for increased learning as a result of using these "motivating" questions, though the gains were less marked than for the review-participation questions. However, because of the differences in question content, no direct comparison was possible in the Yale study as to the placement of corresponding questions before versus after the section of film to which they pertained.

think of the answer, then asking another question, and so on. In the "Review" with "Feedback" experimental group (Group 5: see Table 1), the trainees were told the correct answer after nine seconds had elapsed, followed by an additional two-second pause before proceeding with further questions or instruction.

In the "preview"-participation treatments, the first participation session came before the beginning of the film, while in the "review" treatments the last participation session came after the end of the film. Alternation of film sections and participation, or question-and-answer (Q & A) sessions, may be schematized as follows for the "preview" and "review" conditions:

"Preview" Condition

1st Q & A session	1st section of film	2nd Q & A session	2nd section of film	3rd Q & A session	3rd section of film	4th Q & A session	4th section of film	post-film test

"Review" Condition

1st section of film	1st Q & A session	2nd section of film	2nd Q & A session	3rd section of film	3rd Q & A session	4th section of film	4th Q & A session	post-film test

All experimental groups were given a 30-item test just after the completion of the film instruction. The 30 questions were subdivided by the experimenters into two parallel sets of 15 questions each, designated as Form I and Form II.[3] The question-and-answer participation sessions covered only the points used in Form I (except for one experimental group, Group 3b, which was given practice on only the Form II questions). Thus, all participation groups were given practice in answering questions on <u>half</u> the points covered by the final test, with no practice on the other half. Not all subjects per flight were scored. The score for each flight was composed of the mean score of a random sample of 25 subjects.

[3]Questions in Form I and Form II were grouped into the same four participation sessions as in Experiment I, whether or not the participation session came before or after the section of film to which the questions referred. The testing procedure was very similar to that used in Experiment I, differences occurring primarily in the wording of the instructions when such changes were necessary as a result of the different experimental conditions. All subjects were given 11 seconds to answer each of the 30 questions on the final test. (See Chapter 18 for other details of test content and scoring procedure.)

Experimental Design

The experimental treatments and the arbitrary treatment numbers used to designate them (cf. Table 1, infra) were as follows:

Control Conditions

Group 1: "No Film." Subjects in this control group were simply given the final test, without seeing the film. The scores of the flights constituting this group served as a base-line for determining how much improvement in learning was produced by seeing the film.

Group 2: "Film Only." Subjects in this control group saw the film straight through without participation procedures, and then took the standard test. This control served as a base-line for determining the contributions of the participation procedures to learning of the film content.

Experimental Conditions

Group 3: "Preview Participation." Subjects in this group were asked either the half of the final test questions which constituted Form I or the half which constituted Form II during the course of four "preview" participation periods. Each period of four or five questions immediately preceded the section of film to which the questions applied. The trainees were given eleven seconds to think about each question; they were not told the correct answers directly, but were told at the start of each session that the answers could be found in the next section of the film. (Group 3 was divided into two sub-groups, 3a and 3b, which differed only in that Group 3a was given practice on questions in Form I, while Group 3b was given practice on questions in Form II.)

Group 4: "Review Participation--No Feedback." These subjects practiced on half the final test questions (Form I) during four participation sessions, each session following the section of film that contained the answers to the questions. For Group 4 there was no "feedback" or correction; the subjects were given eleven seconds to try to recall and/or rehearse the answer to each question, and were not told the correct answer during the practice session. (This group corresponds to Groups 8 and 9 in Chapter 18.)

Group 5: "Review Participation With Feedback of the Correct Answer." The procedure for this group was the same as that for Group 4 except that the subjects were given only nine seconds to recall and/or rehearse the answer to each question; after this period they were told the correct answer. (This group corresponds to Groups 10 and 11 in Chapter 18.)

Group 6: A supplementary group, discussed below.

Additional Comparisons

"Test Versus No Test." Half the flights treated under each condition (Groups 2, 3, 4, and 5) were told that they would be tested on the film contents at the end of the viewing. The other flights in each group were not told this.

"Rote Versus Meaningful Learning." In order to investigate the nature of the learning increment effected by covert participation, the wording of approximately half the practice questions for any one experimental group was altered (cf. earlier discussion of the rationale for this procedure). The 15 questions asked during the practice periods were divided into sub-classes of eight and seven questions (designated as Sub-class A and Sub-class B, respectively). The participation questions given to two of the four flights treated under each experimental condition had the wordings of Sub-class A changed (designated Sub-class A'), while the wording of items in Sub-class B remained unchanged. The other two flights per experimental condition were exposed to the opposite situation (i.e., the re-phrased Sub-class was B'). The wording on the final test was identical for all classes; wording changes were introduced only on the questions asked during the practice periods.

Table 1 gives a tabular summary of the main experimental treatments. The supplementary group (Group 6) shown in the table was also tested in order to obtain some preliminary evidence on the question of how much information, in terms of immediate recall of specific factual material, would be imparted by just the question and answer with "feedback" procedures without any film. Using the Form I questions, the four additional flights constituting Group 6 were given the question and answer session alone (with covert participation and KCR), with no film shown. All the questions were given in one session, two of the flights having been told that they would be tested, and the other two not having been told to expect a test.[4]

[4]The question here is whether the contributions of the film and the participation sessions both significantly affected the final learning level for the participation groups—or, for immediate recall of items practiced during the participation sessions, how much the film contributes if the question-and-answer session is used. That is, what is the difference in the level of learning attained when the answers to be learned are rehearsed during question-and-answer ("participation") sessions interspersed between sections of the film (as in Group 5), versus when the answers to be learned are rehearsed during question-and-answer sessions presented without the use of the film (Group 6)? If there were no difference, and if immediate-recall verbal learning were all that was required, then it may be that under many circumstances a well-designed and conducted question-and-answer session could be equally as effective and considerably less expensive than a film. Although these conditions may not often obtain in practice, it was felt worthwhile to run Group 6 for the purpose of making an exploratory comparison with Groups 1, 2, and 5.

TABLE 1

Experimental Treatments
(Numbers in parentheses indicate the
number of flights tested for each condition)

	Control Groups		Experimental Groups				
			Film with Preview (covert) participation; no KCR		Film with (covert) Review participation (Form I practiced)		No Film (covert) participation exercise only (with KCR)
	No Film	Film Only	Form I pract.	Form II pract.	No feedback (No-KCR)	With feedback (KCR)	
Treatment Number:	1	2	3a	3b	4	5	6
Test Announced ("T")	(4)	(2)	(2)	(2)	(2)	(2)	(2)
Test not Announced ("NT")	--	(2)	(2)	(2)	(2)	(2)	(2)

Results

Range of Performance

Table 2 shows that the "base-line" score obtained by the group of flights that were simply given the test without seeing the film (control condition 1) was an average of 15.4 per cent of the total possible score.[5] The review-participation-with-"feedback" group (Group 5) obtained the highest mean score of any group, namely 72.3 per cent of the total possible score on the items practiced (Form I items). On these items, the flights that simply saw the film straight through without any participation procedure obtained an average score of 58.2 per cent. Thus there was considerable room for the various experimental conditions to show improvement over the straight film condition.

Test Announced Versus Test Not Announced

As in Chapter 18, no significant difference was found—contrary to the hopes of the experimenters—in the test scores of those subjects told in advance that they would be tested after viewing the film, and of those not told. For those told that

[5] As in Chapter 18, the mean score for flights within the No-Film Control group is the result of pooling mean scores on Forms I and II.

they would be tested, the mean per cent score of items correct (averaged for groups 2, 3a, 3b, 4, and 5 for all 30 test items) was 56.6 per cent. The corresponding score for the subjects not told that they would be tested was 55.8 per cent. The difference, although in the predicted direction, is quite unreliable ($t < 1$). In consequence of this result, no opportunity was given to compare results for groups shown to differ effectively in motivation, and the flights in the test-announced and test-not-announced categories were treated together in analyzing the data.

TABLE 2

Mean Percent of Correct Answers for the Five Main Experimental Groups*
(Means for items included in the participation sessions are underlined)

	Control Groups		Experimental Groups			
			Preview (covert) participation; no feedback		Review (covert) participation (Form I practiced)	
	No Film	Film Only	Form I pract.	Form II pract.	No feedback (No KCR)	With feedback (KCR)
Treatment Number:	1	2	3a	3b	4	5
Form I	15.37	58.21	<u>60.67</u>	58.92	<u>55.72</u>	<u>72.27</u>
Form II	15.80	48.19	49.57	<u>59.34</u>	45.89	47.23

*Based on mean scores (maximum possible score of 90) shown in Supplementary Table 1.

"Review" Participation Versus Film Shown with No Participation Questions

With "Feedback" of Correct Answers (Group 5). The difference of 14.1 per cent between the mean score of 58.2 per cent for the control group and the mean score of 72.3 per cent for Group 5 is reliable at the three per cent level (see Table 3).[6] When a *t*-test is computed for the second-order differences between

[6] Reference to Supplementary Table 1 at the end of this experiment reveals that variability among flights within the experimental conditions is quite large. It is probable that factors such as intelligence differences are responsible, but unfortunately it was not possible to obtain adequate ability or intelligence-test data on the subjects in this experiment. It is possible, however, to compensate somewhat for large flight-to-flight variance by comparing mean second-order difference scores, for two experimental conditions, where each second-order difference (within flights of one experimental condition) is the difference between the "non-practiced" items score (e.g., Form II) and the "practiced" items score (e.g., Form I). A *t*-test can then be used to determine the frequency with which the two mean-difference (cont.)

Form I and Form II scores, the difference between Groups 5 and 2 is reliable at better than the one per cent level. The results for Group 5 thus confirm the finding of significant improvement in learning of practiced material as a result of the "review" participation with "feedback" procedure (in Chapter 18).

TABLE 3

Tests of Differences between Straight Film Viewing (Group 2) and Review-Participation Conditions (Groups 5 and 4) (Difference scores are for mean percentages of correct answers)

	Inter-group differences (for Form I items only)			Second-order differences (Form I minus Form II)		
	Diff.	t	P	Diff.	t	P
Group 5 (KCR) minus Group 2	14.06	2.56	.03	15.02	7.55	.01
Group 4 (No KCR) minus Group 2	-2.49	1.0	--	-0.19	1.0	--

Without "Feedback" (Group 4). The difference between test means on the practiced items for the film-only control group and the review-participation, "no-feedback" group was slight and non-significant. This is true for a t-test of the difference between the means, and also for the mean differences between Form I and Form II scores for the two conditions (cf. Tables 2 and 3). These results may be compared with those obtained in Chapter 18 for the corresponding "covert"-participation groups. In that instance, some evidence was obtained of a slight gain due to the addition of "covert, no-feedback" procedures, but the effect was even moderately reliable only for the lower-ability subjects.

6 (cont.) scores would be obtained if they were members of the same universe. Such tests of second-order differences are made in the analysis of the results of this experiment, where applicable, in addition to the inter-group first-order-difference comparisons. Regardless of whether first- or second-order differences are employed, t-tests for comparing any two groups in this experiment can be computed either for unpaired measures or for measures paired on the basis of replication, etc.—i.e., line for line in Supplementary Table 1. With four flights per group, the former procedure, of course, gives 6 d/f, the latter 3 d/f. The former procedure (6 d/f) was used for t-tests reported in the text, unless paired measures are specified. P-values for t-tests are given for one tail of the t-distribution.

Comparison of Groups 4 and 5. (Effect of "feedback" of correct answer for review-participation procedures). When the mean score of 72.3 per cent for Group 5 (with "feedback" of the correct answer) is compared directly with the mean score of 55.7 per cent for Group 4 (no-"feedback"), the difference between the means (16.6 per cent) is reliable at about the three per cent level ($t = 2.33$). When the comparison is based on the differences between the Form I and Form II scores, the difference between means is significant at better than the one per cent level ($t = 6.98$). The value of the "feedback" procedure of giving trainees the correct answer after they have attempted to answer each question themselves—shown in the study reported in Chapter 18—is thus confirmed by the data of this experiment.

Effect of the "Preview" Participation Procedure (Groups 3a and 3b)

Comparison with the Film-Only Control Group (cf. Table 2). Separate comparisons were made for Group 3a (which was asked the Form I questions during the "preview" participation sessions) and Group 3b (for which the Form II questions were used).

Effects on "Practiced" Questions (those asked during the participation sessions). For Group 3b, to which the Form II questions were presented in "preview" rehearsal, the learning level attained on the Form II items (59.3 per cent) was significantly greater than that attained on these items by the control group (48.2 per cent) which had no participation sessions ($P < .02$). However, for Group 3a (in which Form I questions were asked preceding each section of the film) the Form I mean score (60.7 per cent) was not reliably greater than the mean score for the film-only control group (58.2 per cent).

The evidence for gains in learning resulting from the use of "preview"-question procedures is thus clear only for one of the two sets of questions. Examination of the question-by-question results further indicated that the significant mean-difference between Group 3b and Group 2 was attributable to the effects of only four or five out of the 15 Form II questions. These few questions appear to have been the only ones that succeeded in "alerting" the audience to get information from the film that would otherwise have been missed. Examination of the content of these questions by several judges failed, however, to reveal any consistent characteristics of content or form differentiating the questions which "alerted" the viewers from the other questions on either Form I or Form II.

Comparisons for Questions not Covered in the Participation Sessions. Comparisons between Group 3a and 3b scores and no-film control-group scores failed to show any appreciable differences on the "non-practiced" items. The means on Form I were 58.9 per cent correct answers for Group 3b and 58.2 per cent for Group 2; for Form II the means were 49.6 per cent (Group 3a) and 48.2 per cent (Group 2). Thus, there is no evidence that the "preview"-question procedure used in this experiment had any effect, either favorable or adverse, on the learning of the other material in succeeding sections of the film, to which the question did not direct the trainees' attention.[7]

Comparison of "Preview"- versus "Review"-Participation Procedures. Two different comparisons can be made: (1) Group 3a ("preview" participation on Form I, without "feedback" of correct answers) versus Group 4 ("review" participation using the same questions, also without "feedback"); and (2) Group 3a versus Group 5 ("preview" participation of Form I questions with "feedback"). The first of the comparisons (no "feedback" with either participation procedure) yields no significant difference between the mean test scores obtained on Form I. For the second comparison the difference of 11.6 per cent (Condition 5 minus Condition 3a) is reliable at about the two per cent level ($t = 2.85$). The second-order difference of 13.9 per cent between mean differences for Form I and Form II is also reliable, in this case at better than the one per cent level ($t = 6.87$).

This experiment provides no evidence, then, for a superiority of one participation procedure over another, unless "feedback" is added to the "review"-question technique, in which case "review" participation appears to be distinctly superior. The evidence is incomplete, however, since a similar comparison is not available for the other set of questions (Form II).[8]

Results for Group 6 (question-and-answer session with no film)

This analysis is confined to the scores on "practiced" items—i.e., questions covered by the question-and-answer sessions. These were the Form I questions of the final test. The supplementary group of four flights, given question-and-answer sessions without seeing the film (Group 6), obtained a mean Form I test score of 67.9 per cent. The difference between this score and the corresponding score of 72.3 per cent for Group 5 (the flights that were given similar participation sessions between sections of the film) is not very reliable ($t = 1.0$, 6 d/f; or using paired scores, 3 d/f, $t = 1.61$, $P \doteq .10$).

When the results for Group 6 are compared on the test items practiced with those for Group 2 (the film-only group), a moderately reliable difference of 9.7 per cent is found between the means of 67.9 per cent for Group 6 and 58.2 per cent for

---

[7]This finding may be compared with some results of an earlier study (Lumsdaine, May, and Hadsell, 1947, 1958) in which the giving of written pre-film tests covering half of the points included in the post-film test was shown to raise somewhat the final scores on the "covered" items, but at the expense of lowering slightly the average scores on the "non-covered" items.

[8]A comparison of Group 5 with Group 3b would have been desirable, especially since it was Group 3b rather than 3a that differed from the film-only group. However, this comparison was not considered legitimate because different participation questions were used—Form I for Group 5, Form II for Group 3b—for which the film-only levels differed markedly.

Group 2 (t = 2.16, P = .04 with 6 d/f; t = 1.81, P = .09, with 3 d/f). For the other material, not covered in the Group 6 procedure, the film group, of course, scored much higher (48.2 per cent) than the mean for Group 6 (20.1 per cent).

Results for Practice with Rephrased Questions

The purpose of this comparison is to ascertain whether effects of the practice given during participation sessions for Group 5, shown previously to result in significant gains over the film-only group, are obtained when the wording of the practice items differs from that of the test items. The comparison was made by obtaining for each flight in Group 5 a sub-score based on test answers to the Form I items that had been rephrased for the participation session of that flight. Similar scores were obtained for Group 6.

The mean scores, in terms of percentage of correct answers, were as follows (see also Supplementary Table 2, at the end of this experiment).

	Group 5	Group 6
Sub-scores for rephrased items:	72.0%	66.0%
Sub-scores for non-rephrased items:	72.4%	69.8%
Difference:	0.4%	3.8%

The difference between rephrased and non-rephrased items is not reliable for either Group 5 or Group 6, or for the two in combination, and thus gives no evidence that rephrasing the items for the practice sessions reduced their effectiveness. The superiority of Group 5 over Group 2 for rephrased items was tested by comparing the percentage scores for rephrased items with the total percentage scores for the Group 2 flights. (The measures used for Group 2 were the total Form I percentage scores; these were compared directly with the rephrased-item Group 5 scores since for the latter group one sub-set of items was reworded for half the flights and the remainder for the other half of the flights.) The mean difference of 14.2 per cent in favor of Group 5 was reliable at the four per cent level (t = 2.26, 6 d/f).

Interpretation and Conclusions

Military Replication of Results Obtained with Civilian Subjects

In terms of the measures used in this experiment, covert-participation procedures seem to have similar effects in producing increments in learning for Air Force basic trainees and civilian high school students.

1. In the present experiment, as in the high school experiment, announcement or lack of announcement of a test to follow the film had no significant effect on the extent of learning from the film and participation procedures. A possible reason for the absence of a "test-anticipation effect" is that the intrinsic level of interest

and the corresponding level of motivation to learn the film's content material was so high that a test threat was unable to heighten attentiveness further.

2. As in the high school experiment, post-film ("review") questions, with covert responses and "feedback" of correct answers, produced significant gains over the learning resulting from the "film-only" control condition. The same participation procedures without "feedback" produced no significant difference from the straight film showing. The importance of the confirmation of the effectiveness of covert participation with "feedback" lies in the likelihood that considerable improvements in learning can be accomplished under the simple filming and film-utilization conditions of covert practice. No special classroom materials such as answer booklets are necessary for covert participation, and the questions and correct answers presumably can be incorporated directly into the film at relatively small cost.

Meaningfulness of the Learning Increments

Analysis of the test scores obtained with reworded questions provides no reason to believe that the learning resulting from participation procedures is purely "rote" rather than "meaningful" in nature.

Effectiveness of "Preview"-Question Participation

The results of this experiment are not definitive on the efficacy of "preview" questions. With Form I questions, "preview" procedures produced no significant gains, but with Form II questions a significant but relatively small improvement occurred. It had been suggested that questions asked prior to the section of film containing the answers might "alert" the audience to the pertinent aspects of the following material, and, by thus spotlighting the crucial points, increase the probability of their being learned. When the trainee is asked a series of "preview" questions, however, he then has to remember something of the questions in order to have an increased chance of finding the correct answers. A series of questions on unfamiliar subject matter may be difficult for the trainee to remember, especially if he is not "set" to remember them. Even when the question is remembered, the correct answer may not be selected as readily as it is in the case of the "feedback" treatments.

Comparison of the scores for questions in Form I and Form II for the "film-only" group indicates that Form II was significantly more difficult than Form I. A question-by-question examination of the two forms, however, revealed no consistent differences either in the nature of the questions themselves or in the answers to the questions given in the film. Consequently, no particular characteristics of the questions or of the answers, on the basis of which a greater or lesser tendency to "alert" the viewers depends, can be suggested.

Effect of Question-and-Answer Session Alone

For the items that were practiced, the use of question-and-answer procedures without the film resulted in a level of learning falling between that attained by the "film-only" control group and that resulting from the film plus "review" participation with "feedback."

There is, of course, considerable material presented in the film that is not learned when the question-and-answer session is used without showing the film. Whether the instructional effectiveness of a question-and-answer session without a film or other instructional presentation would hold up if the attempt were made to cover a far more extensive set of points of information, such as the set covered by the film, is problematical. Nor is it known whether the question-and-answer procedure results in as lasting an effect as would be produced by viewing the film or by participation procedures with the film showing.

Summary

1. Increases in the amount learned by Air Force trainees from a training film (on defense against atomic attack) were investigated when question-and-answer "participation" sessions were inserted between sections of the film. In all cases, participation was of the "covert" form, that is, the trainees were given an interval (of about 11 seconds) to think about each question and try to answer it mentally.

2. The results are, in general, compatible with those of a previous experiment that utilized civilian students. In addition to replicating previous comparisons of covert-participation conditions in a military setting, this experiment involved the investigation of some new conditions of covert participation.

3. When "review" questions were asked in participation sessions, referring to the preceding section of the film, and with the correct answer given by the instructor after the trainees had thought about each question, the post-film test scores showed a significant increase in learning as compared with the scores of the "film-only" condition.

4. The "covert"- or mental-participation sessions were not found to be effective, however, when the correct answers were not supplied by the instructor after the students had responded.

5. Incomplete but suggestive evidence was obtained that "preview"-participation questions (preceding the relevant sections of film) may result in some increase in learning as compared with a straight film showing.

6. The fact that rephrasing the practice questions on the final test did not affect the trainees' scores suggests that the learning increment produced by participation consists of "meaningful" rather than "rote" learning.

SUPPLEMENTARY TABLE 1

"Flight" Mean and Grand-Mean Scores, for Scores on Form I and Form II (maximum score of 90 points in each case). Underlined Values are the Test Means for Items Included in the Participation Sessions

		Control Groups		Experimental Groups				
				"Preview" Participation (Covert Practice) No "Feedback"		"Review" Participation (Covert Practice) (Form I Practiced)		No Film; Form I Questions and (Covert) Answers Only
		No Film	Film Only	(Form I practiced)	(Form II practiced)	No "Feedback"	"Feedback" of Correct Answers	
Treatment Number		1	2	3a	3b	4	5	6
	Test Form							
Test Announced	Form I	14.88	56.80	50.96	55.48	63.72	73.16	63.96
	Form II	17.16	45.96	41.40	54.80	51.36	49.32	21.84
	Form I	16.00	54.48	55.72	52.24	38.40	58.04	60.68
	Form II	15.48	46.92	45.08	53.84	33.32	38.64	18.60
Test Not Announced	Form I	11.72	56.32	55.80	59.08	45.92	60.76	55.64
	Form II	11.40	45.20	45.84	58.44	37.56	39.04	15.64
	Form I	12.72	41.96	55.92	45.32	52.56	68.20	64.08
	Form II	12.84	35.40	46.12	46.56	42.96	43.04	16.24
Grand-Mean Scores:	Form I	13.83	52.39	54.60	53.03	50.15	65.04	61.09
	Form II	14.22	43.37	44.61	53.41	41.30	42.51	18.08

SUPPLEMENTARY TABLE 2

"Flight" Means for Rephrased and
Identical Sub-sets of Form I Questions

Flight	Set	R or i	Number of items	II	V	VI
1	A	R	8	30.20	35.84	28.92
	B	i	7	26.60	37.32	35.04
2	A	i	8	28.44	28.20	30.44
	B	R	7	26.04	29.60	30.24
3	A	R	8	30.64	28.96	26.36
	B	i	7	25.68	31.80	29.28
4	A	i	8	22.52	32.92	30.84
		R	7	19.44	35.28	33.24

Experiment II

The evidence obtained from the study with high school students (Chapter 18) suggests that practice is the major factor in the effectiveness of active-review procedures; i.e., that active review is effective primarily because it makes further practice more likely on the items in the films that are practiced during active review (student participation). For reasons indicated below, it seems desirable to replicate this experiment under lower-motivation conditions.

Motivational Factors

Selective Learning: It is possible that one of the effects of active review (student participation) is to establish a learning "set" by the use of questions during the review sessions (cf. May and Lumsdaine, 1958, Chapter 7). Thus if the questions during the review sessions deal, for example, with minute details, the learner may be "set" to look for and implicitly practice material of a similar nature during subsequent sections of the film, and learning of this kind of material (e.g., fact, minute detail) might be superior to the learning of material dissimilar in nature (e.g., general principles).

General Attention: Active-review (student-participation) effects may lie in the procedure itself—breaks in the film, questions and answers, etc.—and the procedures may act simply to get the learner to pay more attention to the film-showing, and thus produce greater learning. If this is the case, then one would expect greater learning due to active review on all material presented in the film, and not on certain classes of material (e.g., fact and minute detail versus general principles).

These latter (general attention) factors would be much more likely to affect the level of learning under general conditions of low motivation rather than high motivation. In Experiment I, the general condition was one of quite high motivation. President Truman had just recently announced that Russia had successfully exploded an atomic bomb. The film used, Pattern for Survival, was an excellently produced account of how to defend oneself against an atomic attack. The purpose of Experiment II was to attempt to ascertain whether under conditions of lower intrinsic motivation, active review (student participation) would provide significant motivational as well as practice effects. To this end, the film materials selected for study were deliberately chosen to be of quite low intrinsic interest. Two main hypotheses and one secondary hypothesis were tested:

1. It was hypothesized, based on the assumption that increased motivation leads to greater implicit practice, that under conditions of an already existing high general motivation, the effects of active review will tend to be limited to practice effects. This may be the case because, when general motivation is high, the addition of active review probably does not yield sufficient increments in motivation to produce gains on non-practiced items.

2. If there are some motivational effects resulting from the active-review procedure, it would be expected that these would most likely be revealed under conditions of generally low motivation. It is therefore predicted that under this general condition there is the greatest likelihood of gains on non-practiced items. This could be predicted because there is opportunity to increase implicit practice during the film where the base-line level of such implicit practice is low enough to allow room for improvement. This low level of implicit practice would presumably accompany a generally low level of motivation, of the sort generated by an intrinsically uninteresting film, and without the students' having the extrinsic motivation of being told they would be tested.

3. As stated earlier, two of the possible effects of active review are a general attention factor and a selective learning factor. If a selective learning factor is operating, then it would be expected that the active review would establish a learning "set" to look for and practice implicitly material in the film that is similar to the material practiced during the active review. This would predict greater gains on non-practiced test items that are similar to the test items presented during active review, versus the non-practiced test items that are not of the same type as those practiced. (Even this type of effect, however, might be revealed only under conditions of generally low motivation, since under high motivation additional instructional increments might not be revealed in superior performance.)

Method

Film

The film used was entitled, "Airplanes Change the World Map" (distributed by Encyclopedia Brittanica Films). This film traces the history of world map concepts from ancient times to the Air Age and runs for approximately ten minutes (see Supplementary Table 1 at end of chapter). This film material is somewhat difficult and is presented in a static fashion. Inspection of the film convinced the experimenters that, in general, the intrinsic motivation level could safely be considered to be low.

Manipulation of Motivation

In general, levels of motivation can be manipulated by either intrinsic or extrinsic methods. "Intrinsic" motivation is that which results from the film itself; i.e., interesting content, interesting presentation, or both. Extrinsic motivation is that which results from conditions not inherent in the film; e.g., instructions. For this experiment, rather than attempting to produce a condition of low motivation by extrinsic conditions while using a film with high intrinsic motivation, the experimenters used a film of low intrinsic motivation and attempted to raise motivation, for some sub-groups of students, by means of instructions—telling students they would be tested ("high" motivation) versus not telling them ("low" motivation).

Although there is conflicting evidence as to how readily differential levels of motivation can be produced by instructions, there are data that indicate that, at least in some experiments, motivation can be manipulated by instructions. For example, data obtained both by Hovland, Lumsdaine, and Sheffield (1949, Chapter 9) and by Lumsdaine, Sulzer and Kopstein (cf. Chapter 17, this volume) showed—unlike the results of the audience participation studies reported in Chapter 18 and Experiment I of this chapter—that the announcement of a test in advance of the film-showing did produce better learning than no announcement of a test. The lack of demonstrable effects for the test-no-test announcement in the two latter studies was ascribed in part to already high motivation in high school students because the test situation is so familiar to these students. A further reason for the lack of effect on subjects at Sampson Air Force Base (Experiment I) was that the variance obtained with these subjects was so great that any effects would be likely to be masked. Furthermore, with generally high intrinsic motivation due to an intrinsically very interesting subject matter, any additional increment achieved by test announcement may in any case well be negligible. (When a film provides a person with information on how to protect himself against atomic attack, telling him that he will be tested on the content of the film may do very little in motivating him further to attend more carefully!)

Active-Review Procedures

Overt review procedures with feedback (KCR) were used in this experiment. (Responses were practiced by writing them; subjects were told, after they tried to answer each question, what the correct answer should be.) It will be recalled that this method proved to be highly effective for increasing learning on the material practiced, in Chapter 18.

Since the present experiment (Experiment II) was concerned primarily with the motivational effects of active review, and therefore with the performance on non-practiced questions, only 10 questions out of a total 35 were used as review questions. [This proportion also had the possible advantage that it paralleled more nearly that used in an earlier experiment by Lumsdaine, May, and Hadsell (1947, 1958), in which significant effects of "motivating" questions had been found.] There were three review sessions. The first review session (following the first section of the film) had four questions; the second had four questions; and the third (at the end of the film) had two questions.

Review Question Content Sets

In order to test the speculations concerning selective learning, it was necessary to select questions for the review sessions that might serve to establish a "set". The questions on the test were divided into those emphasizing fact and those emphasizing principle. "Fact" questions were defined as questions of specific detail such as places, names, dates, and distance relationships, whose answers were to be found directly in either the auditory or visual parts of the film. "Principle" questions were defined as questions of a non-specific nature, the answers to which were not to be found directly in the film and which thus had to be answered inferentially from the film content. All the questions practiced during the active review sessions were fact questions with the intention of establishing an implicit fact "set." ("Set" as used here means any technique—explicit or implicit—that informs the subjects about the responses they are to learn.) Illustrative items of the fact and principle types are presented below.

"Fact" Questions:

1. In what year did Mercator draw a map of the earth on a flat surface?

2. Name two fifteenth-century centers of commerce.

"Principle" Questions:

1. There are two situations in which a great-circle route drawn on a Mercator map will be a straight line. Where are they?

2. Why is there no single two-dimensional map which is best for all purposes?

Measuring Instrument

A test of 35 items was used, broken down as follows: 10 "fact" items practiced, 20 "fact" items not practiced, and five "principle" items not practiced. All items were of the open-end type, answerable by one word or a short phrase. A total of six points could be obtained on each item, with partial scores of 2, 3, or 4 possible on several items. Thus the maximum possible total score was 210, with partial maximum obtainable scores of 60, 120, and 30, respectively, for the three above-named categories of items.

Subjects

Subjects were 993 trainees at Sampson Air Force Base, in the seventh day of basic training. Classes used were made up of half-flights (20 to 32 S's). The experimental sessions occurred between 9:30 and 11:00 A.M. and from 2:00 to 4:30 P.M., and the average experimental session lasted about 45 minutes. All the experimental sessions took place in the same room, so that, with the exception of the time differences, the environmental stimuli were presumed to be constant for all groups.

Design

Diagramatically, the design was as follows (numbers in the cells indicate the numbers of classes tested under each condition):

Motivation	No Review	Active Review	No Film
High (Test announced)	9	8	4
Low (Test not announced)	9	8	

Eighteen classes were thus shown the film without any review. These classes will be referred to as "no-review" groups (NR). Half of these classes were run under conditions of high motivation (H), and half under conditions of low motivation (L), manipulated as indicated above. The "no-review" groups served as the control groups in this experiment. Sixteen classes were shown the film with active review (AR). Half of these classes were run under conditions of high motivation and the other half under low motivation.

Results

Initial Ability

The mean total test score of the four classes tested without viewing the film was 19.51. There were several questions on which no correct answers were given. This very low mean indicates an extremely low level of knowledge of the film material prior to the viewing of the film.

Learning Produced by Film Alone

When a comparison is made between the groups that received the film without review and the groups tested without seeing the film, large significant differences are obtained. A mean of 60.37 (total score) is obtained by the control groups (film only), whereas the no-film group obtained a mean of 19.51. Although the amount learned by the group that viewed the film is still relatively small (29 per cent of the total possible score), the increase over the no-film group indicates that viewing the film resulted in the learning of the material presented.

Differences Due to Motivation Conditions

Insofar as the predictions in this experiment are dependent upon the existence of conditions of high and low motivation, all the subsequent analyses and conclusions are dependent upon evidence that motivational differences occurred. The two no-review groups were used to test for a differential performance due to differences in motivation. (If the active-review procedures had any motivational effects, these effects might confound differences due to the effects of motivation derived from the instructions alone.)

This analysis revealed a significantly superior performance (on the non-practiced test items) for the test-announced (high-motivation) group: (See Table 4.)

TABLE 4

Comparison of Mean Score* for the Test-Announced ("High Motivation") and No-Test-Announced ("Low Motivation") No-Review (Participation) Condition

Groups Compared	M	df	t	P (one tail)
"Test" (NR)	38.38			
		16	2.24	<.01
"No Test" (NR)	33.93			

*Based on data for the non-practiced items

Thus, as anticipated, under conditions of generally low motivation, the announcement of a test in advance can demonstrably improve performance.

The Effects of Active Review on Learning of Review (Practiced) Items

An analysis based on the ten items reviewed during the active-review session reveals large significant differences in favor of those groups that received the active review—both for the [Test (AR) group and for the No-Test (AR)] group. These results are presented in Table 5.

TABLE 5

Comparison of Mean Score* for the Active-Review and No-Review Groups on the Review Items, under Test-Announced and Test-Not-Announced Conditions

Groups Compared	M	df	t	P (one tail)
No Test (NR)	13.67	15	18.90	<.001
No Test (AR)	33.17			
Test (NR)	14.32			
		15	10.95	<.001
Test (AR)	34.14			

*Based on the ten practiced test items

These results are consistent with the previous finding that active review with provision of knowledge of the correct response (feedback) produces increases in learning on items practiced during the active-review (audience-participation) sessions.

The Effects of Active Review on the Non-Review Items

The major concern in this study was, of course, to determine whether the active-review procedure not only has direct practice effects (as previously demonstrated) but also, under certain conditions, can serve to motivate the learner (and thereby increase his performance on the items not reviewed). As will be recalled, it was hypothesized that if such motivational effects do occur, they would be more likely to occur under conditions of generally low motivation.

These hypotheses are supported by the results of this analysis. Examination of Table 6 reveals that under the "test-announced" conditions of generally high motivation there is no significant gain in the non-review items for the groups which had active review over the no-review groups. By contrast, under conditions of low motivation (no test) a significant gain does occur. Here, under low motivation, performance on the non-review items is significantly better for the active-review groups than for the no-active-review groups.

TABLE 6

Comparison of Mean Scores on Non-practiced Items for the Active-Review Versus No-Review Groups Under Conditions of High-Motivation (Test Announced) and Low-Motivation (Test Not Announced)

Groups Compared	M	df	t	P (one tail)
High-Motivation				
Test (AR)	38.38	15	.01	--
Test (NR)	38.42			
Low-Motivation				
No Test (AR)	37.36	15	1.85	.05
No Test (NR)	33.93			

The Effects of Intelligence on Test Performance

Although the results presented in the preceding section reveal significant differences in directions predicted, it is desirable to adjust for the effects of intelligence (as measured by the Air Force Qualifying Test), since a large correlation (r = .65) was obtained between AFQT scores and test performance. This large correlation suggests that intelligence enters as a factor influencing the results. Analysis of co-variance was used to correct for the effects of intelligence, and all the analyses presented in the preceding sections were re-analyzed on this basis. The results of these analyses are as follows:

The Effects of the Motivational Instructions. Table 7 presents the results of the co-variance analysis between low and high motivation (test-not-announced versus test announced), groups with the correction for intelligence.

TABLE 7

Analysis of Co-variance for Effects of Motivating Instruction—the "Test" (NR) Versus "No-Test" (NR) Conditions

Source of Variation	SS	df	MS	F
Between ("test" vs "no-test")	47.01	1	47.01	4.77*
Within (error)	147.95	15	9.86	
Total	194.96	16		

*F is significant at the .05 level with one and 15 degrees of freedom.

This analysis reveals that intelligence has a very slight effect upon the results in this comparison. The t-test (Table 4) yielded a t = 2.24, whereas the t based on the analysis of co-variance is 2.18. Both tests reveal significantly higher performance in favor of the groups that were given instructions designed to produce high motivation.

The Effects of Active Review on the Non-Practiced Items. Tables 8-A and 8-B, respectively, present the co-variance-analysis evidence on these effects for low-motivation conditions—comparison between "No Test" (NR) versus "No Test" (AR)—and for high-motivation conditions—"Test" (NR) versus "Test" (AR).

TABLE 8

A. Analysis of Co-variance Between "NR" and "AR" Under Low-Motivation ("Test") Conditions

Source of variation	SS	df	MS	F
Between (AR—NR)	58.08	1	58.08	6.85**
Within (error)	118.61	14	8.47	
Total	176.69	15		

B. Analysis of Co-variance Between "NR" and "AR" under High-Motivation ("No-Test") Conditions

Source of variation	SS	df	MS	F
Between (AR—NR)	12.51	1	12.51	1.02*
Within (error)	170.38	14	12.17	
Total	182.89	15		

*F not significant **F Significant at Ca .02

This analysis of co-variance tends further to substantiate the hypotheses tested in this study. The significance of the differences increased in both comparisons when the co-variance technique is employed. However, the same trend of differences still exists—i.e., significant differences between the No-Test (NR) and No-Test (AR) conditions, and no significant differences between the Test (NR) and Test (AR) conditions.[9]

General-Attention Versus Selective-Learning Effects

Two types of effects, or factors, were postulated for the operation of the active-review procedures (aside from direct practice effects). These were (1) a "general-attention" motivational factor, and (2) a "selective-learning" factor. It was reasoned that if active review (AR) produced a "set," i.e., selective learning, then the AR groups' performance on non-practiced items should differ from that of the NR group for only those items which were similar to the review items (here, the fact items). On the other hand, if active review did not wholly function selectively (i.e., the general-attention factor was operative), then the differences should be in the same direction for both the "fact" items and "principle" items. Finally, the effects of both factors—but particularily that of the general-attention factor—would be expected to be more marked under conditions of low extrinsic motivation (no-test-announcement).

Tables 9-A and 9-B, respectively, present the comparisons for the effects of AR versus NR based on the "principle" items alone, for Low-Motivation conditions (A) and for High-Motivation (B).

TABLE 9

A. Analysis of Co-variance Between "NR" and "AR", Based on 5 "Principle" Items: for Low-Motivation ("No Test Announced") Condition

Source of variation	SS	df	MS	F
Between (AR—NR)	2.71	1	2.71	2.95*
Within (error)	12.89	14	.92	
Total	15.60	15		

*F is not significant: $\underline{t}$ based on F ratio: $\underline{t} = 1.72$
$\underline{P} = .05$ (one tail)

B. Analysis of Co-variance Between "NR" and "AR", Based on 5 "Principle" Items: for High-Motivation ("Test-Announced") Condition

Source of variation	SS	df	MS	F
Between (AR—NR)	1.18	1	1.18	1.35*
Within (error)	12.27	14	.87	
Total	13.45	15		

*F is not significant. $\underline{t}$ based on F ratio $\underline{t} = 1.16$
$\underline{P}$= not significant (0.1, 1 tail)

9See Supplementary Note at end of chapter.

Although the differences are neither as great nor as reliable as the differences presented in the previous comparison, it is evident, at least, that the differences are in the same direction.[10] Thus the motivational effects of the active-review procedure on non-covered items seem to be general-attention effects. However, this conclusion is, at best, tentative and suggestive—due, for one thing, to the relatively few items on which this analysis is based.

Discussion of Results

1. This experiment helps to clarify the apparent discrepant results in previous studies on the effects of informing subjects about to see an instructional film that they will be tested afterwards. Hovland, Lumsdaine, and Sheffield (1949) found that such test announcements can enhance learning from films. In the study reported in Chapter 18, no such gain due to test announcement was obtained. On the basis of all three studies (Chapters 18 and 19). it would now seem that when interest level in the material presented is generally low, the prior announcement of a test to follow the presentation can enhance motivation to learn, but that when the presentation already has a high interest value to viewers, such prior announcement of a test may add little or nothing to the attention and consequent learning level. Thus the effort to manipulate the overall level of motivation, in order to investigate its interaction with the relative contributions of practice and motivation in student participation, failed with a highly interesting film (Chapter 18) and succeeded with a very dull one (Experiment II above).

2. Experiment II tends to confirm and support the results of previous studies concerning the effects of student-participation procedures. When these procedures are used, considerable gains in learning take place, particularly in the information rehearsed ("practiced") during the student-participation practice sessions.

3. This experiment contributes to our understanding of the relative contributions of practice-of-specific-content and of motivation-to-learn from films to the audience-participation effect. Hovland, Lumsdaine, and Sheffield (1949) pointed out that the audience-participation effect could have been due either to the motivating effect of the procedure, its effect on the subjects' attention to the contents of the film, or to the added practice of the correct responses that the procedure guaranteed. In Chapter 18, Michael and Maccoby found that with the highly interesting film that was used (on civil defense against nuclear attack) the learning gains from student-participation exercises, though substantial, appeared to be confined entirely to the added practice that the procedure provides. The results of Experiment II parallel those of Michael and Maccoby in failing to find significant "motivational" gains under conditions of generally high motivation. However, they also demonstrate that under conditions of low motivation (i.e., using a generally dull film without the incentive of announcing that students will be tested on it), gains, though not very large ones, can be obtained on the non-reviewed material. Clearly, under conditions of generally low motivation, some "indirect" gains due to student-participation procedures can be obtained. There is, further, suggestive evidence

[10] See Supplementary Note at end of chapter.

that, at least in part, these latter gains reflect a "motivational" factor of enhanced general attentiveness rather than solely a selective—attention factor.

Supplementary Note

[The hypotheses with which the analysis of indirect effects on non-practiced material is concerned can actually be stated in two related ways. First, we can say that under conditions of low extrinsic motivation (such that effects would not be obscured by a "ceiling" because individuals already have a level of motivation that is difficult to augment by special instructions) two things should be true: (1) active-response procedures will have motivational carry-over and/or other indirect effects on material not specifically practiced (as well as having direct-practice effects); and (2) this effect will, however, be reduced, and may even disappear, when motivation is raised from its otherwise low extrinsic level by such special instructions as the announcement to students that they will be tested. In other words, this form of the hypothesis predicts indirect (e.g., motivational) effects when extrinsic motivation is high, on the one hand, but small or negligible indirect effects on non-practiced material if motivation is sufficiently raised. In the extreme case, the latter aspect of this hypothesis (absence of effects under high motivation) represents an experimentally untestable proposition, since here the <u>experimental</u> hypothesis is in fact a "null" hypothesis, which is statistically unprovable. A second form of the general hypothesis is (for this reason and possibly for others) therefore preferable. This form of the hypothesis states merely that the indirect effects of active response procedures will be <u>greater</u> under low motivation conditions than under high motivation conditions. (It thus does not demand the further proposition that, under a given, relatively low motivational level, the indirect effects of the active response procedure on non-practiced material will necessarily be <u>zero</u>, and thus avoids the methodological impasse just noted.)

To test the hypothesis in the latter, preferred, form requires an examination—in analysis of variance terms—of the interaction variance between the effects of active response and motivation level. In what the Editor considers to be more straightforward terms, its proof requires demonstration of a significant second-order difference—namely, showing that the difference between the "AR-NR" scores on non-practiced material is greater under low motivation conditions than is the corresponding "AR-NR" difference under high motivation conditions. [In this experiment, "NR" represents no-review as against active response (AR). In other experimental contexts, the appropriate first-order difference could be "AR-PR" (the superiority in effect of active review over passive review)—cf. Hovland, Lumsdaine, and Sheffield, 1949, Ch. 9, and discussions of active review comparisons reported herein.]

The direct testing of this second-order difference in "motivational" effects of active review on non-practiced material under conditions of high versus low extrinsic motivation can, of course, be made either in terms of the raw class means presented in the third column of data in Supplementary Table 2 (<u>infra</u>), or in terms of derived values for which an attempt is made to reduce error variance on the basis of relevant correlated variables. The latter, in this case, obviously include intelligence level (utilized in the co-variance analyses presented in Tables 7, 8 and 9) and also, perhaps, other information—including the performance of the various class groups on the practiced material. The use of this second predictor variable is, of course, complicated by the fact that one pair of experimental groups (AR versus NR) differed in procedure (and, also significantly, in resulting scores on the practiced items); but some use of this information would nevertheless seem possible for one of the two stages in the second-order-difference comparison if it is considered that $(AR - NR)_{Hi} - (AR - NR)_{Lo} = (AR_{Hi} - AR_{Lo}) - (NR_{Hi} - NR_{Lo})$— with "Hi" and "Lo" standing, respectively, for "high motivation" (test announced) and for "low motivation" (test not announced).

The advantage in statistical power of considering the data in paired classes, or even (without using AR scores as a control variable) in quartets of classes identified on the basis of AFQT ranks, will, of course, depend on whether the correlation between the stratifying and criterion variables is sufficient to more than offset the loss in degrees of freedom. This loss in degrees of freedom would arise primarily, of course, from the difference in number of d/f appropriate to paired versus unpaired comparisons, but also, in this case, is affected slightly by the fact that use of pairs or quartets involves reducing by one class the number of independently-considered class observations in the "NR" groups (which contain nine classes each as against only 8 classes per group for the "AR" condition in this experiment). Further use of such a stratifying variable as AFQT might also be made, with recourse to the original scores of individual students within classes, by some form of partial or more complete "equating" procedure (as well as using the mean AFQT scores as a basis for pairing or for co-variance analysis).

In the present case, without access to the original data for individual S's, an attempt to investigate the significance of the relevant second-order difference must be limited to the class means presented in Supplementary Table 2. Utilizing only the AFQT scores as the stratifying variable, and arbitrarily discarding the ninth class in each of the NR groups to provide four quartets of classes for the four experimental groups (quartets being identified simply in terms of corresponding rank order of AFQT means within groups), yields eight second-order differences (-11.10, 1.76, 1.23, 16.18, 2.88, 9.68, -1.39, 0.18), having a mean of 2.43. Wilcoxon's signed-ranks test for paired replicates for these six positive and two negative differences does not attain significance; *t* for these differences, is less than unity.

A somewhat analogous logic applies to the comparison of "general-attention" and "selective-learning" factors. Since both factors could be operative, with the former especially expected to be a more potent factor under conditions of low extrinsic motivation, we can ask whether the "transfer" effect of active response (AR versus NR) to non-practiced "principle" items occurs (1) under high motivation conditions (Table 9-B); (2) whether it occurs under low motivation conditions (Table 9-A); and (3) whether, in any case, the difference in transfer effect to the "principle" items is greater under low motivation than under high motivation conditions. If both the "general-attention" and "selective-learning" factors were operative, we should predict that all three of these questions could be answered in the affirmative; whereas, for example, if only the "selective-learning" factor were operative, the third (second-order) difference might not obtain, and if only the "general-attention" factor were operative, we should expect to find the second-order difference (3), and to find no difference (even though the lack of it would not be provable) with the highly motivated learners (1). —Ed.]

SUPPLEMENTARY TABLE 1

Editing Schedule for Experimental Versions—Experiment II

The Airplane Changes Our World Map*

Content	Cumulative Running Time
1. 1st section: beginning "our earth is. . ." ending "Behaim's globe, and others".	2'.91"
2. 2nd section: beginning "since there are. . ." ending "to and from these basins".	7'.31"
3. 3rd section: beginning "We still think. . ." ending "the surface of the earth".	10'.46"

*specified in minutes (') and seconds (") from designated reference points.

SUPPLEMENTARY TABLE 2

Basic Data Showing Class Means for Each Treatment and Class Means on Different Types of Items

	Class*	Total (35 items)	Practiced ("Fact") Items	Non-Practiced Items	"Principle" Items	AFQT Scores
High Motivation ("Test") Group; NR	11	59.48	12.67	35.79	4.48	60.81
	12	63.63	15.36	38.00	5.08	65.80
	13	63.42	13.58	41.00	6.21	54.79
	14	69.00	16.80	39.50	4.93	63.16
	15	50.48	12.75	29.33	2.96	50.87
	16	70.79	18.00	41.79	5.62	64.16
	17	66.00	12.88	44.00	6.54	62.16
	18	60.94	15.09	35.53	3.97	50.62
	19	64.08	13.40	40.48	6.72	59.08
					M =	59.05
High Motivation ("Test") Group; AR	21	79.54	34.88	34.17	5.12	46.83
	22	89.31	34.86	42.86	7.07	65.34
	23	82.61	35.74	34.87	5.61	56.60
	24	96.92	40.08	43.92	4.77	60.38
	25	99.57	39.43	45.25	5.14	62.71
	26	70.91	27.39	33.35	4.17	37.86
	27	79.74	31.89	35.96	5.15	58.59
	28	76.22	29.90	36.94	4.87	51.90
					M =	54.91
Low Motivation ("No-test") Group; NR	31	53.14	11.78	31.50	4.32	55.03
	32	53.87	12.87	31.75	3.50	58.25
	33	57.58	11.58	35.75	4.81	50.74
	34	70.42	16.35	41.61	6.10	66.77
	35	56.87	15.00	32.29	4.03	57.22
	36	63.65	15.58	36.31	5.08	58.58
	37	53.15	13.35	30.80	4.60	51.05
	38	63.68	15.95	37.09	4.73	64.23
	39	46.52	11.35	28.26	1.74	48.78
					M =	56.73

SUPPLEMENTARY TABLE 2 — Continued

	Class*	Total (35 items)	Practiced ("Fact") Items	Non-Practiced Items	"Principle" Items	AFQT Scores
Low Motivation ("No-test") Group; AR	41	77.00	32.87	34.25	3.28	56.50
	42	75.15	29.22	35.37	5.67	61.26
	43	82.84	34.69	37.12	4.85	55.00
	44	80.11	34.07	34.25	4.43	53.79
	45	76.93	32.46	33.75	5.14	49.39
	46	83.74	31.63	39.89	5.85	57.81
	47	89.38	35.07	42.31	6.00	58.24
	48	88.04	35.37	41.96	6.21	57.89
					M =	56.23
No Film	51	17.83	3.26	--	1.39	52.09
	52	23.19	5.19	--	2.38	58.11
	53	19.67	4.20	--	.93	63.07
	54	17.35	2.65	--	.46	49.50
					M =	55.69

*The first number in the column indicates the treatment; 1 denotes condition Test (NR); 2, condition Test (AR); 3, condition No Test (NR); 4, condition No Test (AR); and 5 the No-Film condition. The second number is the class number assigned to that class within each treatment.

PART III

STUDIES ON LEARNING OF PAIRED-ASSOCIATES MATERIAL

INTRODUCTION TO PART III

Each of the studies reported in Part III deals (with one partial exception in Chapter 23) with the learning of some form of paired-associate materials and all deal with the learning of relatively direct associations between stimulus and response elements which can be rather simply and precisely described. A number of conceptual and experimental parallels exist between these studies and experiments dealing with other forms of learning tasks and materials (including those reported in Parts I and II of this volume). These parallel relationships suggest strongly that the implications of these studies are by no means limited solely to the acquisition of paired-associates material. This is not to deprecate the practical importance of paired-associates learning which, though sometimes regarded as representing an artificial task, actually enters into a remarkably large proportion of the material which has to be learned in both academic and technical training—e.g., the pairing of names and faces, commodities and prices, components and nomenclature, symptoms and causes, and foreign language vocabulary equivalents, to name a few. At the same time, it is clear that broader implications from paired-associates experiments, transcending the paired-associates context as such, can increase not only their practical interest but their theoretical significance.

In addition to its concern with theoretically relevant variables, the study reported by Kopstein and Roshal in Chapter 20 has at least limited "face importance" in that the learning materials used were foreign-language equivalents—objects and their Russian names. The study is further of interest in that one of the variables investigated was pictorial depiction, so that its relevance to programmed audio-visual materials may be somewhat more readily perceived than in the case of studies which deal only with verbal-symbolic material in abstract form. This study was an outgrowth of an earlier investigation by Lumsdaine (1949, 1950, 1958) in which, unlike the Kopstein and Roshal experiment, verbal and pictorial representations for paired associates were used both in the stimulus and response positions.

The question of pictorial versus verbal representation in the experiment reported by Kopstein and Roshal in Chapter 20 was only one aspect of a more general concern with representational factors, and the experimenters also manipulated temporal relationships between stimulus and response terms, as well as the use of contextual verbal material. The latter variation (verbal context) suggests, superficially at least, a bridge between the stark format of paired-associate material and the format of so-called "connected discourse" material. (As a matter of fact, many elements in the latter actually reduce to paired-associate equivalents, at least insofar as can be inferred from the material on which the learner is tested in many objective examinations.)

Chapter 21, prepared by J. O. Cook, differs rather markedly in style and conception from most of the symposium papers prepared for this volume. Aside from being presented in rather informal style, it has a somewhat autobiographical flavor which, with only a moderate amount of blue penciling, the editor has decided to let stand. The story it has to tell of the behind-the-scenes, generally unpublicized aspects of research could, of course, be paralleled in many other instances (including the sequences of events that led to the preparation of each of the other papers in this volume). This is an account of the informal etiology of hypotheses, and of some of the workaday details of practical experimentation, adventitious historical accidents, humor, pathos—and, often, blood, sweat, and tears—that lie behind the production of findings that are then reported in relatively staid, objective accounts in technical papers such as the majority of those presented in this volume. As the author notes, such details are generally left undocumented. However, perhaps in at least one illustrative case their presentation—as contrasted with merely the factual description of procedures finally adopted, tables of data, and end-products of after-the-fact interpretation—may prove instructive.

The account in Chapter 21 derives primarily from a series of experiments by J. O. Cook, T. S. Kendler, and H. H. Kendler, the results of which, as the author notes, led directly or indirectly to a series of subsequent studies, including those which Cook reports in Chapter 22 and the study reported by Angell and Lumsdaine in Chapter 24. In Chapter 22, Cook presents an attempt at an integrative summary of response analysis in paired-associates learning. The experiments he reviews in relation to the analysis of factors involved in paired-associates learning draw on the experiments conducted in the New York University series (Chapter 21), subsequent results of experiments by Cook and his collaborators that are detailed in Chapter 22, and on the work of other investigators, including the Kopstein and Roshal study (Chapter 20). The issue of the functions served by "prompting" versus "confirmation" techniques dealt with here is of considerable relevance to the design of auto-instructional programs, and forms one of the experimental antecedents for current studies centering around the use of prompting and of "feedback" or "reinforcement" panels in such programs.

A second antecedent experiment which was, in effect, also concerned with this prompting-confirmation issue (although it was initially described in different terms) is the experiment reported by Briggs in Chapter 23. This chapter is based on experiments performed in conjunction with author's projects at the AFPTRC Maintenance Laboratory, and was carried out under contract at Tulane University by A.L. Irion. The relation of this experiment to subsequently developed concepts and experimentation on auto-instructional programs has been noted previously by Carr (1959) and others. The Briggs and Irion study is further of interest in that it represents one of the few attempts, albeit an unsuccessful one, to provide experimental evidence bearing directly on the question of what is currently termed "vanishing" or "fading" of cue support. Such a procedure is a deduction from S-R learning theory (cf. Lumsdaine, 1959b) and is one intuitively followed in some fashion by most teachers. Curiously, there has been some difficulty getting its efficacy to show up in clean-cut fashion in controlled studies. (The not wholly consistent results of Maccoby and Sheffield, as summarized in Chapter 5 for procedural learning, perhaps represent the most clear-cut demonstration to date.)

Use of a partial degree of prompting (though not progressive decrease in amount of cue support provided the learner) is the basic issue behind the study of Angell and Lumsdaine reported in Chapter 24. This study furnishes moderately convincing support for the general proposition that some intermediate degree of prompting should be superior to complete prompting. It grew directly out of the Kendler and Cook studies summarized in Chapters 21 and 22, and used the same paired-associates materials. Although it did not attempt to replicate Cook's experiments fully, it reproduced one of his conditions—the one used by Cook to represent "prompting," but which, as the authors point out, actually consisted of a mixture of prompted and unprompted test trials. The unprompted trials were actually introduced by Cook as test trials (an interesting illustration of the more general problem in the methodology of science in which the technique of measurement may modify the phenomenon under observation). Angell and Lumsdaine, accordingly, compared a prompting condition in which no unprompted test trials were introduced until fairly late in the course of acquisition, in order to compare this "purer" prompting condition with the mixed condition of prompted and unprompted trials which Cook had employed to contrast with his "confirmation" condition.

The factor of pacing, dealt with by Briggs in Chapter 25 in terms of individually-varied self-pacing versus a fixed pace of presentation imposed by a program-presentation device, is another factor which would seem to have a central bearing on the advantages of individual pacing advocated by proponents of auto-instructional devices. The directness of translation for the results reported by Briggs is, however, somewhat obscured by the fact that in his experiment the total amount of time utilized by the subjects under self-paced and externally paced conditions was kept constant. (Cf. also the experiment and relevant discussion by Fry, 1960.) Nevertheless, the results do help to point up the issue of whether the student is the best "programmer" of his own activities with respect to the rate through which he works with these materials in the course of learning. The interesting possibility is suggested that better results might be achieved by a program-varied gradation of pacing—i.e., from slower pacing at first to faster pacing later on.

The work reported by McGuire in Chapters 26 and 27 was a continuation of the previous series of studies at Yale University by the same author, reported in Chapters 12, 13, and 14. Although the present studies employed paired-associates material, in pictorial-verbal form (learning nomenclature for mechanical parts), they have continuity with McGuire's earlier-reported studies on motor learning in terms of the variables with which both experimental series were concerned. In particular Chapter 26, like Chapter 14, was concerned with the introduction of "motivational" statements within the instructional series and its relation to distribution of practice. Unlike the Chapter 14 study, this experiment also introduced a control for the confounding factor of spacing of practice which accompanied the introduction of motivational statements (by providing equivalent silent pauses for a control group).

The overt-covert response variable studied by Michael and Maccoby (see Chapter 18), as well as in a number of subsequent experiments on auto-instructional material, was one of the two factors introduced in the further experimental analysis reported by McGuire in Chapter 27. Time was also manipulated as a variable, but—unlike some of the later auto-instructional studies—was not confounded as a consequent of the overt-covert factor. Rather, in McGuire's study, the two factors were independently co-varied, and a comparison was made between overt and covert responding under fast versus slow rates of presentation. Also of interest in Chapter 27 is the theoretically oriented analysis of factors (distribution of practice, "feedback," motivation, response generalization, and various deleterious effects, including those associated with interference between participation and "reception" of the presentation) with which the first part of the chapter is concerned.

Chapter 28, by Kanner and Sulzer, is both a direct and indirect descendant of the original phonetic-alphabet study on "audience participation" reported by Hovland, Lumsdaine, and Sheffield (1949, Chapter 9). It employed the same practically oriented paired-associates task, and used similar basic audio-visual materials. However, in terms of the design and experimental variations employed, the experiment reported by Kanner and Sulzer grew more directly out of the Michael and Maccoby experiment (see Chapter 18), since the major experimental variations here employed included a comparison of overt and covert responding, and practice of only a portion of the material versus explicit practice of all of it. (The reversion to paired-associates material from the substantive or "connected-discourse" subject matter employed by Michael and Maccoby suggests a further illustration of the "back to paired-associates" trend chronicled by Cook in Chapter 21.) This later study, conducted by Kanner and presented jointly here by Kanner and Sulzer, introduced the further variation of including both overt oral responding and overt written responding—the former automatically providing oral feedback from other subjects in the class group as in earlier experiments with this subject matter. In it controls were included to assess not only the contribution of active review versus no review, but also (unlike the Michael and Maccoby experiment) to compare the effect of active review with equivalent passive review. The marked advantage of any of the forms of active review—overt or covert, written or oral, etc.—as compared with passive review (which was, in turn, significantly though not grossly superior to the no-review condition) form an interesting contrast with recent studies in which this superiority has not uniformly been revealed. (See further discussion in concluding chapter.)

The study reported by Wulff and Stolurow in Chapter 29 is an interesting example of studies of paired-associates learning which deal with organizational factors having at least suggestive implications for the sequencing and organization of material in auto-instructional programs (as differentiated from specific, solely intra-frame or intra-segment programming techniques). This paper deals with "class-descriptive" cues that differentiate sub-groups of material which a trainee must learn to identify, and contrasts two forms of organization that can be used in sequencing in the instruction. "Class-organization" displays showed all of the items together in relation to one set of features, or differentiating cues, whereas "item organization" displays gave the same information but with all features or characteristics of a given item in the to-be-learned list presented at one time. The latter method was considered to make it theoretically unlikely that the learner could utilize class cues effectively during learning, and hence was predicted to be an inferior method.

The work reported by Wulff and Emeson in the final chapter of this part of the volume (Chapter 30) comprises several small-scale experiments conducted to explore the implications of a general rationale. This rationale relates particularly to difficult training tasks, and involves the dual principle of not only insuring correct practice, but also arranging training conditions so that the practice of correct responses will be appropriate. More explicitly, Wulff and Emeson argue that the determination of what is "appropriate" requires analysis of what are the really difficult aspects of what must be learned, and that this in turn requires not just identification of the required criterion performance but analysis of "steps" that the learner must successively master to get from his initial repertoire to this required terminal behavior. They provide an experimental demonstration with a sample paired-associates task to show how it is possible, when such an analysis has been correctly made, to devise an "engineered" (or, we could say, "programmed") sequence which will result in highly efficient training. An incidental aspect of this study, which is interesting in relation to sometimes over-zealous assumptions that psychologist-generated programs will necessarily surpass less structured conditions of acquisition arranged by the student (cf. also Newman, 1957), is found in the results for a heuristic comparison of a formalized (but "non-engineered") paired-associate training schedule as compared with unstructured self-study.

CHAPTER 20

VERBAL LEARNING EFFICIENCY AS INFLUENCED BY THE MANIPULATION OF REPRESENTATIONAL RESPONSE PROCESSES: PICTORIAL-VERBAL AND TEMPORAL CONTIGUITY FACTORS[1]

Felix F. Kopstein
Sol M. Roshal

Introduction

The present experiment employed paired-associate materials, with Russian words as response terms and familiar objects as stimulus terms. Stimulus terms were presented pictorially as line-drawings of the stimulus objects for some experimental groups, and as English words naming the object for other experimental groups. An additional experimental variable involved the temporal relationship between the presentation of stimulus and response terms, which were presented either simultaneously or in a staggered arrangement in which the stimulus term first appeared, and then, after an interval, was joined by the response term.

Theoretical Considerations

This research was based on the premise that factors which may be presumed to affect the character of implicit, or mediating, or representational responses taking place during verbal learning, either directly or indirectly, can be expected to produce appreciable differences in learning efficiency. The premise rests in turn upon repeated observations that the efficiency of rote verbal learning appears to be

[1]This study, initiated at the Air Force Human Factors Operations Research Laboratory in 1953, was completed by the authors at the Training Aids Research Laboratory, AFPTRC, in 1954-55, and was under the general direction of A.A. Lumsdaine, then TARL Technical Director. Preliminary experimental results of the study were reported at the 1954 and 1955 meetings of the American Psychological Association; these papers were also given limited distributions as mimeographed reports (Kopstein & Roshal, 1954, 1955). The senior author was until recently on the staff of the Training Research Branch of the ARDC Behavioral Sciences Laboratory (formerly Aero-Space Medical Laboratory), and is now in charge of research on auto-instructional methods for the Burroughs Corporation. The junior author, who had coordinate responsibility for the planning and design of the study, is now in private practice in Los Angeles, where he serves as consultant to various industrial and governmental institutions and to several research projects at the University of California, Los Angeles. The senior author prepared the present paper and assumes responsibility for the analysis and interpretation of the data as presented here, except for a few expository changes introduced by the Editor.

quite independent of gross variations in the teaching or study conditions (Wright & Taylor, 1949; Newman, 1957; Kopstein, 1961). Wright and Taylor, as well as Kopstein, found, moreover, that the one factor to which rote verbal learning efficiency could be related was the total practice time per item, which may suggest that the learner's covert behavior is of major significance.

Pictorial Representation

It is reasonable to expect that facilitation of the covert behavior of the learner should contribute to learning efficiency, and a number of experimental findings support that expectation. Roshal (1949), for example, found that the learning of a perceptual-motor task through films becomes more effective as the film approaches a representation of the learner himself performing the task to be learned (i.e., requires less stimulus generalization). In comparing the relative efficacy of pictorial representations of objects and of their verbal equivalents as stimulus-and-response terms in a paired-associate learning task requiring verbal responses. Lumsdaine (1949, 1950, 1958) showed that pictures were more effective than words as stimulus terms, though less effective as response terms. Superiority of pictures over words was also found by Herman, Broussard, and Todd (1951), who worked with a serial list. Wulff and Stolurow (1957: See Chapter 29) obtained confirmation for their hypothesis that training conditions that favored the augmentation of implicit item-descriptive cues by implicit class-descriptive cues would facilitate the mastery of a paired-associate learning task. Other findings of this type could also be cited.

In the present experiment, one method of facilitating covert learning behavior was examined in order to shed further light on the generality of the empirical findings, particularly when different learning materials are used, and under modifying conditions of context, practice, and test. The particular technique used was similar to Lumsdaine's, involving the use of pictures versus words as stimulus terms in a paired-associate learning task. Although he had reliably established the superiority of pictures as stimulus terms, he had done so only for situations in which both stimulus term and response term were meaningful and familiar. There is practical as well as theoretical interest in the question of whether these findings would continue to hold when response terms were meaningless and unfamiliar.

Practical interest derives from a variety of teaching situations ranging from simple vocabulary or nomenclature learning through auto-instructional techniques. Theoretical interest involves two factors that Lumsdaine invoked in accounting for the superiority of pictures. First, he asserted, if the required response in a paired-associate learning task is verbal, the use of a picture in the stimulus position of each pair will interfere less than will a word with an assumed implicit (presumably sub-vocal) practice of the verbal response term. In terms of this factor, an unfamiliar response term requiring a greater degree of implicit practice may well enhance the superiority of pictures—as compared to a familiar response term. Second, Lumsdaine pointed out that a word occupying the stimulus position may be variably interpreted, thus providing less constant stimulus conditions than can be provided by pictures. This factor may operate in two ways. Either the variability of interpretation provides an unstable stimulus situation which can impair learning, or, in the course of variation, an already well-established association with a familiar response term may be hit upon and facilitate learning. However, if the response term is new and unfamiliar, the chances of hitting upon an already established association with the familiar term in the stimulus position are presumably small. The influence of the unstable stimulus condition, however, would not be eliminated. In order to minimize this factor in the present study only such stimulus terms were used as had been empirically shown to be highly unequivocal for the subject population.

Simultaneous versus Staggered Presentation of Terms

In presenting a paired-associate item it is possible to expose stimulus and response term simultaneously, or to expose the stimulus term alone for a while, and only then to expose the response term as well. It was reasoned that in the latter situation a subject could make an implicit provisional response to the stimulus term after that particular item had appeared at least once previously. Then, as soon as the response term appeared, the subject would receive either confirmation or disconfirmation for his tentative response. Confirmed responses should be strengthened or "fixated," while others should be extinguished or inhibited. On this basis, a "staggered" presentation should prove superior to a "simultaneous" presentation. In the present experiment this also might mean that a staggering of stimulus term and response term should emphasize any differences that might exist between picture and word cues. For example, suppose that one category of cues (let us say the pictures) is superior. During the first presentation of the list a relatively larger number of items (associations) will be acquired with picture cues than with word cues. During the second repetition of the list in its picture version a relatively larger number of items will be correctly anticipated, and hence fixated, than in its word version. On further repetitions a relatively greater and greater number of items in the favored version should be confirmed, fixated, and retained in the subject's active response repertoire. The end result, if the assumptions are correct, is a multiplicative rate of item mastery for the favored version over the less favored "good" version.

However, it is conceivable that during the early stages of learning the staggered method of presentation may prove to be relatively ineffective. Being entirely new and unfamiliar, the response terms themselves have to be learned, quite apart from learning to pair them correctly with the stimulus terms. The shorter exposure of the response terms under conditions of staggered presentation may be detrimental to the learning of the response terms themselves and consequently to the mastery of the list.

Verbal Context

As noted earlier, the anticipated efficacy of pictures as stimulus terms was based, in part, on the assumption that the implicit response to a picture is of a different type from that to a word and interferes minimally with implicit responses to the word-response terms. Placing the picture in a verbal context could be presumed to establish a set for implicit verbalization that might modify the character of, or interfere with, the implicit response normally made to the picture. If this were actually the case, some indications of the same interference effects postulated for words would be expected even though pictures were used. The expectation thus was that picture-stimulus terms presented in a sentence context would be less effective than without this context, though perhaps slightly better than word-stimulus terms.

Transfer-testing Conditions

In a pilot study for the present experiment (Kopstein & Roshal, 1954) it had been found that individuals who had been trained with pictures, but were tested with printed English words as cues, performed less well than those who had been trained and tested with pictures. In attempting to account for this discrepancy, it was suggested by the authors that, among other things, the former situation made it necessary for the subject to make a transfer from the picture-stimulus term to the word-stimulus term that had never been shown in association with the response term. This explanation implied that similar results were to be expected whenever the test conditions differed from the training conditions with respect to the type of cue employed. In the present experiment, provisions were made to obtain further information about this matter.[2]

Method

Materials

A list of English words for use as paired associates was selected from the 1,000 most common terms listed in the Thorndike-Lorge Word List by culling those which met criteria of: (a) "objectness," (b) concreteness, (c) familiarity, (d) easy representation in picture form, and (e) avoidance of common homonyms. (However, selection of words with no homonyms turned out to be quite impractical.) The first list of words selected was then submitted to five judges, together with the above criteria. The judges were instructed further to select those words which, in their opinion, met the criteria. Only the words upon which there was unanimous agreement among all judges and the experimenters were retained. At this stage, the list of English words was translated into Russian, and the Cyrillic spelling was approximated as closely as possible in the Latin alphabet. From this list those words were then further selected for actual use whose response term (the Russian word) contained exactly five letters. A final list of eight pairs was retained as follows:

[2] [A study of pictorial and verbal symbols in learning foreign language material was subsequently initiated by Kale (Kale, Grosslight, and McIntyre, 1955). This study also used English-Russian pairs, with somewhat longer lists than in the present study. Both still and motion-picture depiction of the objects or actions represented by the words (nouns or verbs) were employed. In the Kale study, unlike the present one, verbal representation of the printed stimulus term was always employed, even when there was pictorial accompaniment. It was concluded that either still or moving pictures made a positive contribution to the learning to write the foreign words, compared to no pictures—Ed.]

1. Table - Stohl
2. Cup - Chasa
3. Leg - Nagah
4. Chain - Tsyep
5. Cross - Krest
6. Window - Aknoh
7. Hand - Rokah
8. Bridge - Mohst

In order to avoid serial-learning effects, the serial order of the pairs within the list was scrambled according to a table of random numbers for each of 12 repetitions. Sixteen versions of the 12 repetitions of this list, corresponding to 16 cells in the experimental design, were prepared on 16-mm. motion picture film. Those versions which employed pictures showed, for each item, a pictorial representation of the term in the stimulus position, on the left side of any frame, and the Russian response term printed on the right side. The versions employing words had the stimulus term represented in printed instead of pictorial form.[3] (See Figure 1)

Where a verbal setting was desired for either type of cue, an enclosing sentence "A.....is called....." was provided (see C and D in Figure 1).

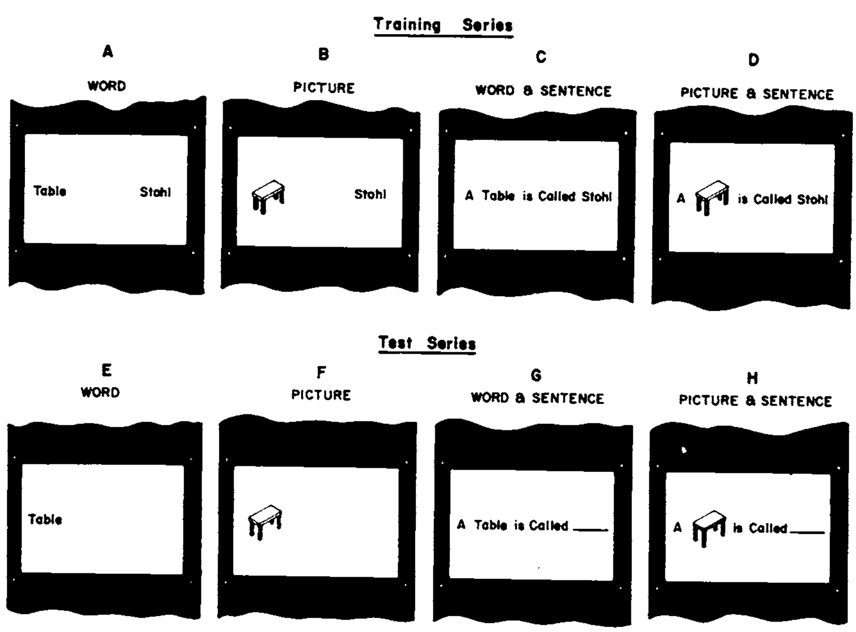


Figure 1. A sample item and how it appeared in different film versions

[3]Whether the printed and pictorial forms of any item represented the same thing to subjects in general was checked on an independent sample. On the whole, disagreement was found to be negligible.

For a staggered display of stimulus and response terms, the stimulus term alone was shown for two seconds (48 film frames at the standard speed of 24 frames per second), and then the response term appeared together with the stimulus term for an additional 1-1/2 seconds (36 film frames). For simultaneous presentation, stimulus and response terms were presented together for 3-1/2 seconds. The test series (shown in the lower half of Figure 1) corresponded exactly to the training series, save for the omission of the response term. In eight of the film versions, however, picture-training series were followed by word-test series and vice versa, although no similar switch occurred with respect to the "contextual" sentence.

Experimental Design

The sampling in this experiment is based on the mean scores of small groups ("classes"). Thus, the df are obtained from the number of groups tested, and not from the individual subjects comprising them. The four factors under consideration, i.e., type of training (verbal versus pictorial S-terms), type of testing, verbal context, and staggered versus simultaneous presentation, were entered in a factorial design. With three observations (test series) on each group, this yielded a 2 x 2 x 2 x 2 x 3 table. This was replicated six times, and, hence, contained a grand total of 287 df.

Subjects

Subjects for the experiment were 2,080 Basic Airmen in the routine testing and counseling stage at an Air Force Basic Training Center. They were given no hint that the experimental procedure was not a regular part of this routine testing. The subjects reported in flights (approximately 60-75 men) and were arbitrarily divided into three groups. Each group (one-third of a flight) was then randomly assigned to one of the 16 treatments (films).[4]

Procedure

The 12 scrambled repetitions of the list of paired associates were assigned as follows: three repetitions (trials) for a first training series, followed by a test trial; three more training trials, and another test trial; and a final, three-trial training series followed by a final test trial.

As noted above, presentation time for each pair was 3-1/2 seconds. (When a staggered presentation was utilized, the response term appeared only for the last 1-1/2 seconds; for the first two seconds the stimulus term alone appeared on the screen.) Six seconds, or 144 frames of film, were allowed for each item in the test series in which the stimulus term only was shown.

[4]In practice, the assignment was completely random across the first two of the six replications, and then the assignment pattern was repeated for the next two replications, and so on.

The experimental sessions were held in a classroom set aside for this purpose. Since during any experimental session, two thirds of a flight was always waiting in an anteroom, and frequently half of this group had already completed the experimental procedures, great care was taken to prevent any communication between the two groups. The seating arrangement in the experimental classroom was such as to insure a good view of the screen, and to make cheating during the test series difficult. Answer booklets, in which the responses were to be recorded in written form, and pencils were provided by proctors.

At the beginning of the experimental session, the proctors gave instructions on how to fill out the cover sheet of the test booklet on which identifying and biographical information was requested. The projector was then started, and kept running at the standard 16-mm. sound-motion-picture speed of 24 frames per second. Further printed instructions and a demonstration of how to record answers in the test booklets appeared on the screen and, simultaneously, a proctor read them aloud. Also shown were a sample training item and a sample test item, each of which corresponded precisely to the particular film version, but neither of which appeared on the experimental list. In the test series, when only the stimulus term was shown on the screen, subjects were asked to write their responses in a test booklet. Test booklets were so designed as to minimize the possibility of copying items answered during a preceding test series. Instructions required that the subjects fold any completed page of the answer booklet all the way to the back. Thus, the possibility of copying from previous test series due to some transparency of the paper was also minimized. After the end of the complete experimental film, test booklets were closed, and the proctors requested the subjects to record their past language experience according to detailed instructions. At the end of the session, subjects left the room and the next group was brought in.

Results

Responses recorded in the test booklets were scored on a simple "right or wrong" basis. An attempt was made to score a portion of the responses with complex partial credits, but it quickly became apparent that information obtainable in this way was identical with that obtainable by the simpler method. Thus, for any one test series, the total possible score was eight, and the sum total possible score for all three test series was 24.

Prior Knowledge

Of the 2,080 men participating in this experiment, approximately one per cent (21 men) indicated some knowledge of Russian. Of these, only one-half, or approximately .5 per cent of all subjects, had a knowledge of Russian that they rated as fair or better. A total of only 96 individuals, or 4.6 per cent, indicated any degree of knowledge of any of the Slavic languages. These individuals were quite evenly distributed throughout the various experimental treatments. Proceeding on the somewhat improbable assumption that all of these 96 subjects would have obtained the perfect score of eight on a test series given prior to any instruction, they would produce a maximum initial mean score of merely .36. Hence, to all intents and purposes, the level of the prior knowledge for the subject population as a whole may be considered zero.

The Effects of Instruction

If the initial level of knowledge is, in fact, taken to have been zero, then this motion-picture training procedure (without regard to versions) could be said to have produced a 47.6 per cent level of mastery. This percentage is based upon the mean number of correct responses obtained by all subjects on the third and last test series, which, of course, represents the highest level of learning achieved.

There was also, as might be expected, a steady improvement from test series to test series ($P < .001$). These increases were regular, and entirely consistent with normal progression along the customary learning curve. Nowhere, even under the conditions found empirically to be the most favorable, was there any indication of ceiling effect due to the restriction of the score and error range. Mean scores for the 16 films ranged from 1.15 on the first test series of the "least good" version to 4.42 on the last test series of the "best" version.

Equivalence of "Picture" and "Word" Test Forms

Before proceeding to an examination of differences in performance as a function of different experimental treatments, it will be of interest to know whether there was any difference due to the test form *per se*. Although the two basic test forms (picture and word) and the conditions under which they were administered corresponded in every respect, except for the symbolic representation of the items, it is possible that the intrinsically different ways of representing the items in the two test forms may have affected performance.

The differences in obtained mean number of correct responses over the three test series between picture-test form and word-test form, however, fell far short of the five per cent level of significance. Thus, without any evidence to the contrary, we may assume that the two forms of the test series did *not*, in themselves, contribute to the difference between picture and word treatment.[5]

[5] Further support for this contention comes from the observation that the relative order of item-difficulty is remarkably stable. Counting the number of correct responses to each item on each of the three test series, ranking them, and then correlating the rank orders, produced these rank-order coefficients: $Rho_{12} = .992$, $rho_{23} = .988$, $rho_{13} = .996$. When such correlations were run between picture and word training treatment on each of the test series, the following coefficients were obtained: Test series 1 $rho_{PW} = .996$, test series 2 $rho_{PW} = .989$, and test series 3 $rho_{PW} = .984$. It would seem, then, that not only did the relative item difficulties fail to change as learning progressed, but also the type of training treatment failed to affect the relative item-difficulty differentially, whatever the stage of learning.

Pictures versus Words in Training

When the performance of classes[6] receiving picture-training is compared with that of classes receiving word-training, it may be seen that the picture-trained classes uniformly give a very significantly greater number of correct responses (see Table 1).

TABLE 1

Mean Number of Correct Responses with PICTURE and WORD Training*

	Test Series			
	1	2	3	All Test Series Combined
Picture Training	1.66	3.11	3.97	8.74
Word Training	1.44	2.78	3.66	7.88
Picture minus Word	.22	.33	.31	.86

*see also Table 5

This occurs about equally at each of the three stages in the progress of learning at which a measure was obtained, and is not confined to any particular stage of progress.

Transfer Between Symbolic Forms and Related Effects

It has been shown in Table 1 (and may be seen in Table 5) that picture and word forms of the test series did not, in themselves, appear to affect performance. One may also ask whether the correspondence of the test series to the training series was of any consequence. This is an entirely different question, but may be answered via a simple re-combination of the data. Instead of classifying the test series into "picture test" and "word test," those test series that do correspond in form to their respective training series are combined into one classification, and those that do not are combined into another. Whereas previously we had a picture-tested and a word-tested mean, we now have a picture-trained-and-picture-tested-plus-word-trained-and word-tested mean and a picture-trained-and-word-tested-plus-word-trained-and-picture-tested mean.

It will be readily apparent that the "same" versus "different" classification in Table 2 corresponds to the interaction between training and testing, which, through a simple re-labeling, now becomes a main source of potential effect (cf. Table 5). This effect, however, also falls far short of the five per cent level of significance.

[6] The term "class" is here used to distinguish the statistical unit of observation (one-third of a flight) from the experimental subject (a single person). A single score, as used in the present analysis, is a class-score, or, in other words, the mean score of one-third of a flight.

TABLE 2

Mean Number of Correct Responses with Test Series Similar and Different From Training Series*

	Test Series			
	1	2	3	All Test Series Combined
Same Test	1.59	2.92	3.84	8.35
Different Test	1.51	2.97	3.79	8.27
Same minus Different	.08	-.05	.05	.08

*see also Table 5

If, on the other hand, this effect is examined in relation to the presence or absence of a contextual sentence, the interaction is significant beyond the five per cent level. While for simple paired-associate-format, test series which differed in form from their training series affected performance adversely (top of Table 3), they actually produced a uniformly larger number of correct responses in the presence of a verbal setting (bottom of Table 3).

TABLE 3

Mean Number of Correct Responses with "Test Same" and "Test Different" As a Function of a Contextual Sentence*

	Test Series			
	1	2	3	All Test Series Combined
A. No Contextual Sentence				
Test Same	1.65	3.03	3.95	8.63
Test Different	1.44	2.84	3.67	7.95
"Same"-"Different"	.21	.19	.28	.68
B. With Contextual Sentence				
Test Same	1.54	2.81	3.72	8.07
Test Different	1.58	3.10	3.91	8.60
"Same"-"Different"	-.04	-.29	-.19	-.53

*see also Table 5

The character of the interaction effect may be more fully appreciated, perhaps, by considering it as a second-order interaction. A breakdown into training-treatment, by test-treatment, by presence-or-absence-of-verbal-setting, and its assessment by means of Duncan's New Multiple-Range Test (Duncan, 1955) shows that the reversal in the lower half of Table 3 is primarily due to the performance of picture-trained and word-tested classes as against that of the picture-trained and picture-tested classes (P <.01). No significant difference in performance is found between picture-trained and picture-tested classes, word-trained and picture-tested classes and word-trained and word-tested classes. In the absence of a verbal setting (cf. upper half of Table 4), we find that picture-trained-and-tested class groups achieve the largest mean number of correct responses, the picture-trained-word-tested groups are next, and then the word-trained-and-tested groups, with the word-trained-picture-tested people last. Only the first rank is significantly different from

TABLE 4

Mean Number of Correct Responses as a Function of Training, Test and Contextual Sentence

	Test Series			
	1	2	3	All Test Series Combined
A. No Contextual Sentence				
Pic. Train. & Test	1.73	3.27	4.21	9.21
Pic. Train.-Word Test	1.57	2.99	3.81	8.37
Word Train.-Pic. Test	1.30	2.69	3.52	7.51
Word Train. & Test	1.57	2.80	3.69	8.06
B. With Contextual Sentence				
Pic. Train. & Test	1.57	2.81	3.69	8.07
Pic. Train.-Word Test	1.76	3.34	4.12	9.22
Word Train.-Pic. Test	1.41	2.87	3.71	7.99
Word Train. & Test	1.51	2.80	3.76	8.07

all others (P <.05); thereafter, significant differences between adjacent achievement ranks cannot be demonstrated. The level of performance attained by the two top-ranking groups in the upper and lower halves of Table 4 is essentially the same. Except for the picture-trained-and-word-tested ones, each of the classes on the lower half of Table 4 performed at about the same level as the word-trained-word-tested classes on the upper half.

A complete analysis of variance is presented in Table 5.

Presentation Methods

A comparison of the two presentation methods (showing stimulus and response term simultaneously, or staggering their appearance) revealed that on the first two test series the simultaneous display produced the larger number of correct responses by a material and highly significant margin. On the third test series, however, the margin is substantially diminished and the difference between the two display conditions is no longer significant, even at the five per cent level. This undoubtedly accounts for the presence of a significant interaction between progress along the learning curve (i.e., number of repeated list presentations) and the presentation factor, and suggests that the curves obtainable with the two presentation conditions may tend to meet or intersect as complete mastery is approached.

TABLE 5

Analysis of Variance of Number of Correct Responses on Three Test Series Under Sixteen Conditions

Source of Variation	df		MS	F
I. Between classes				
"P" & "W" Training (A)	1		5.263	8.461**
"P" & "W" Testing (B)	1		.500	.804
"Like" & "Unlike" Testing (A x B, or B')	1		.055	.088
"Sim." & "Stag." Presentation (C)	1		9.114	14.653**
"Verbal Setting" (D)	1		.014	.022
A x B x D or B' x D	1		3.131	5.034*
All other between treatments & interactions combined	9		.623	1.001
Between treatments		15		
Between classes within treat. (Error I)		80	.622	
Total between classes		95		
II. Within classes (variations in stage of training)				
Progress of Training (P)	2		125.312	2278.400**
P x C	2		.216	3.927*
All other P x Treatments Interactions	28		.064	1.164
P x Treatments		32		
Pooled within treatments Classes x P (Error II)		160	.055	
Total within classes		192		
Grand Total		287		

* p <.05
** p <.01

TABLE 6

Mean Number of Correct Responses with a Simultaneous and a Staggered Presentation of Stimulus and Response Terms

	Test Series			
	1	2	3	All Test Series Combined
Simultaneous	1.76	3.15	3.94	8.85
Staggered	1.35	2.74	3.69	7.78
Simultaneous minus Staggered	.41	.41	.25	1.07

Discussion

The results of this experiment tend to lend credence to the premise that factors affecting representational response processes can produce substantial differences in learning efficiency. The premise is borne out most clearly by the unquestionable and uniform superiority of the pictures over words as stimulus terms. These results give strong support to Lumsdaine's (1949, 1950, 1958) experimental results concerning the efficacy of pictures as stimuli. The support is less clear for one of the postulates which he proffers as underlying the hypothesis. He asserts that a picture in the stimulus position of a pair will interfere less with implicit practice of the verbal response term than an equivalent word. This position is clearly based on contiguity considerations. However, identical predictions can be made from a position such as Gibson's (1940) still prevalent generalization-differentiation theory. The difference would lie only in Gibson's emphasis on the mutual confusibility of the stimulus terms as the crucial factor and her neglect of the response terms. Either view, however, must accept the existence of the implicit, representational response in order to account for the obtained results. Lumsdaine includes this notion in his theory. Gibson does not deal with this question explicitly; however, two of her principles are relevant here: (a) pictures prove superior because they are better differentiated from each other; (b) all differentiation is due to differential reinforcement. Of course, in the experiment reported here, pictures did not receive a greater amount of differential reinforcement than did words. The only other possibility—i.e., unequal amounts of pre-differentiation for pictures and words—seems to the writer too far-fetched to merit serious consideration. Since no differential reinforcement can be reasonably assumed for the raw stimulus patterns, and since they had been shown independently to be equivalent, only a presumed difference in generalization gradients around some mediating, representational responses can account for the results.

A further appreciation of the operation of the representational response processes can be obtained by considering the significant second-order interaction between the type of stimulus term employed in training and testing, and the presence or absence of contextual material. It must be noted, first of all, that the highest levels of achievement obtained either with or without a contextual sentence were about equal. The difference lies in the type of treatment producing the highest level of achievement. The shift from highest achievement by picture-trained-and-tested subjects under "no context" conditions to picture-trained-and-word-tested subjects under "context" conditions seems to account for the significance of the interaction. This shift, as well as the grouping of the remaining treatments, becomes quite meaningful if four hypothetical factors are considered. Unfortunately their quantitative assay is not possible, and so the interpretation is not conclusively verifiable.

An overall ordering of the treatments, as shown in the upper and the lower halves of Table 4 in terms of the level of achievement produced by them, can be predicted as follows. The hypothetical joint operation of these four factors is summarized in Table 7. The first and most important factor presumed to affect the order is the form of the stimulus-term (here called "representational response"), i.e., either picture or word. This factor is assumed to operate positively for picture and negatively for word. The second factor is the transfer of training to the test situation, which is here assumed to operate either negatively or not at all. The third factor has to do with a "training set" toward verbal forms of representational response, presumably induced by a contextual sentence and operative during training. This factor is thought to be absent when no contextual sentence appears and otherwise to operate either positively or negatively. The fourth and final factor—"test set"—corresponds to the third one, but is here assumed to operate, as shown, in the test situation. In the absence of information about the relative weights of four

TABLE 7

Summary of the Operation of Four Hypothetical Factors

Treatment		Factors			
Training (Stimulus term)	Test	(1) Form of "rR" (Stimulus term)	(2) Transfer	(3) Tng. Set	(4) Test Set
No Context					
Picture	Picture	+	0	0	0
Picture	Word	+	-	0	0
Word	Picture	-	-	0	0
Word	Word	-	0	0	0
Context					
Picture	Picture	+	0	-	-
Picture	Word	+	-	+	+
Word	Picture	-	-	0	-
Word	Word	-	0	0	0

factors (but assuming each to have some weight in the direction indicated) one would, on the basis of Table 7, clearly anticipate a ranking of 1, 2, 4, 3 in the upper half of Table 7, and a ranking of 2, 1, 4, 3 in the lower half. This is almost exactly the actual ranking that occurs quite uniformly in all three test series in Table 4. However, in both halves of the table only the position of first ranks (as against others) are supported by differences of statistical significance. It would be idle to speculate whether the remaining rank orders are spurious, or due to unequal weighting of the factors.

Although the major hypothesis concerning the effects of eliciting different types of representational responses has been well borne out, the secondary hypothesis concerning the enhancement of this effect with a staggered presentation of stimulus and response terms has received no support [since, though the latter differed significantly from simultaneous presentation, it did not interact with the pictorial-verbal factors: cf. Table 5]. It will be recalled that the hypothesis was based upon a number of assumptions, of which the major one is not regarded as tenable. Contrary to expectations, the simultaneous form of presentation was unquestionably superior to the staggered form. Hence the essential condition for the occurrence of the enhancement effect did not exist. The fact that it did not—that more or less the opposite results were obtained—is noteworthy. Unfortunately, an unambiguous interpretation is hampered by the confounding of a synchronization factor with a length-of-exposure factor under the staggered presentation.

Despite the confounding, however, it would appear that the possibility of the synchronization factor as the sole responsible one can be ruled out. The synchronized presentation of stimulus and response terms is operationally equivalent, or very similar, to the procedure which Cook and Kendler (1956) dubbed "prompting," and which Irion and Briggs (1957) called a "quiz" mode of operation in their Subject-Matter Trainer (cf. Chapter 21-23). Cook and Kendler found prompting to be the superior condition for learning, at least up to about the 50 per cent learning level. Irion and Briggs found the quiz mode to be superior even when relatively advanced levels of learning were achieved. Finally, Cook (1958) carried learning to the 100 per cent level, yet prompting continued to be the superior learning condition. Thus it appears to this writer that when synchronization is the sole factor, contiguity of the stimulus-and-response-term presentation leads to more rapid learning at all levels of mastery. This was not true in the present case, as demonstrated by the significant interaction between the temporal presentation factor and the progress of learning. Near the 50 per cent learning level the staggered presentation was no longer significantly different from the simultaneous presentation (cf. discussion by Angell and Lumsdaine, Chapter 24).

There is reason to expect that the superiority of one form of presentation (simultaneous versus staggered) would be primarily a function of the length of the response-term exposure and of the point on the learning curve at which the measurement is made. Since each of the relatively complex response terms was unfamiliar, it may be assumed that initially a relatively non-differentiated implicit response will have been made to it (or that the stimulus characteristics of the implicit response will not yet have been differentiated). Clearly, a successful pairing of the implicit response to a stimulus term (or its stimulus properties) with the implicit response to the response term would have but a low probability under these conditions. A prolonged exposure of the response term should permit longer implicit practice and thus establishment of a well-differentiated implicit response (Wright & Taylor, 1949, and Kopstein, 1961). In turn, a well-differentiated response should enhance the probability of a correct pairing.

These considerations apply chiefly to the early stages of learning. After several trials the implicit responses to the response terms will have become fairly well differentiated. Hence, the length of time for which the response term is actually visible would be far less crucial, since, under the staggered treatment, paired items can now be covertly practiced before response terms become visible to the subjects. Also, the staggered presentation follows the pattern of the test situation quite closely, so that with continuing practice a good deal of positive transfer might be expected.

Summary

An experiment was conducted to shed further light on the effects of manipulating "representational response processes" on learning efficiency. The influence of modifying conditions of context, practice, and test were explored. Stimulus terms in an eight-item paired-associate list were presented either as pictures, or as their equivalent words. The response terms were Russian words. Stimulus and response terms were presented either simultaneously, or in a staggered fashion with an abbreviated exposure of the response term. One-half of all subjects were shown the stimulus and response terms embedded in a contextual sentence. Both picture-trained and word-trained subjects were tested either with pictures, or with words. Subjects were 2,080 Basic Airmen. The use of pictures as stimulus terms proved to be uniformly more effective than the use of the equivalent words. Neither test form alone nor context alone contributed any significant variance. There was a high degree of transfer from picture training to word testing, and vice versa. Although the presence of a contextual sentence appeared to be a hindrance for picture-trained-and-tested subjects, it facilitated transfer for those trained with pictures and tested with words to such an extent that they were able to perform at about the same level as those subjects who were trained and tested with pictures and without any contextual sentence. Contrary to expectations, a simultaneous presentation of stimulus and response terms was more effective than a staggered presentation. However, the superiority of the simultaneous form of presentation diminished at about the 50 per cent mastery level. Reasons are suggested for assuming that the duration of the response-term exposure could be a factor responsible for this result, rather than solely the synchrony of stimulus and response terms. It is concluded that the character of the representational response processes is a major factor in verbal learning efficiency.

CHAPTER 21

FROM AUDIENCE PARTICIPATION TO PAIRED-ASSOCIATE LEARNING[1]

John Oliver Cook

From the face that psychology presents to the world in the pages of some of its technical journals, one might suspect that experiments are generally done solely to test carefully conceived theories. One rigorously deduces a prediction from a theory, and then by some unspecified feat of ingenuity he designs a neat, tight, sharp-edged experiment whose results will either confirm or disconfirm the theoretical implications. What the usual experimental report leaves out is, however, frequently as important, and certainly as humanly interesting, as what it includes. No mention is made of the brain-racking search for appropriate experimental materials, or the hours spent adjusting the apparatus in the hope that it will hold together long enough to run all the subjects. Nothing is said about the bright idea contributed by a secretary, the unforeseen bugs that turn up and have to be exterminated, the tedium of compiling the data, or the resolution of misgivings about the statistical procedure. The arguments between collaborators are passed over in dignified silence. In fact, the whole intellectual soil and climate that nurtures the research is almost entirely neglected. It is hard to find out from published reports why one line of research comes to a dead stop and another one continues, why one idea trails off into nothing at all while another one flourishes. But these things are worth knowing, and it is in large part with them that this paper will be concerned.

The research discussed in this chapter began in 1951 when a contract was let by the Human Resources Research Laboratories (HRRL) of the U. S. Air Force,

---

[1]The studies with which this chapter is mainly concerned were conducted at New York University under Air Force Contracts AF 33(038)-23596 and AF 18(600)-42, directed by H. H. Kendler and monitored by A. A. Lumsdaine, as part of the program of the Human Resources Research Laboratories, and, later, the AFPTRC Training Aids Research Laboratory, during the period circa 1951-1954. In addition to original contractor's typed reports, most of the studies were reported in mimeographed research memoranda or laboratory notes that have had only quite limited circulation. Most of these are noted in the chapter. In addition, the results of one major experiment and associated theory were published by J. O. Cook and T. S. Kendler (1956). The present paper was prepared by Dr. Cook, with the review and concurrence of Dr. Kendler. Dr. Cook is now Associate Professor of Psychology at North Carolina State College, Raleigh, N. C. Appendix material derived from previous reports on these studies as supplied by the author has been abridged and rearranged by the editor.

to the Psychological Research Center of New York University. After we had exposed ourselves to the problems of audio-visual learning long enough to become acquainted with them, but, we hoped, not so long that we had acquired all the prevailing psychological sets toward those problems, Dr. Howard H. Kendler, D. Tracy S. Kendler and I undertook to examine learning theory with a view to teasing out what implications we could from the Hull-Spence point of view that would have some relevance in the area of learning from films. Our ideas, in the form of experimental sketches, were fitted into the format of Drive, Cue, Response, and Reward, a format that has subsequently been utilized also by Neal E. Miller and collaborators (1957) as a conceptual framework for organizing research findings on audio-visual media and discussing their implications. We set to work independently producing ideas. We wrote them down, passed them around, and discussed them. Sometimes an idea by one of us would be hailed by the other two, sometimes it would simply be deplored, and at other times we would argue about it. Sometimes a new idea would appear out of the argument over an old one. It very soon became obvious to all three of us that we were not merely deducing predictions from S-R reinforcement theory. Sometimes the idea came first, and then we tried to deduce it from theory. In other cases it was apparent that we were using S-R reinforcement theory merely as a launching pad from which our ideas took off; Hullian theory suggested an idea or prediction, but we couldn't deduce the idea from the theory. But besides serving as a launching pad for new ideas, reinforcement theory also provided us with the language with which we communicated with each other. The residue of these discussions we eventually translated into a report (Kendler, Kendler, and Cook, 1951) published in mimeographed form by HRRL and which, though not widely available, has been cited in a number of instances in the published literature.

After we had completed this effort, a series of meetings took place between the three of us from N. Y. U. and Dr. Arthur A. Lumsdaine, Dr. Sol M. Roshal, and other members of the staff of HRRL. These meetings were concerned with the problem of selecting an experiment for us to do out of those we had proposed in our first paper. One that was mutually decided upon was "An investigation of the interacting effects of repetition and audience participation on learning from films."[2]

[2]In this experiment (Kendler, Cook, and Kendler, 1956), the main interest centered in the interaction between participation conditions and number of repetitions of review. The main experimental design was a 2 x 2 x 3 layout, the variables being: 1) sex; 2) number of repetitions of review (1, 2, or 3 repetitions); and 3) type of instructions to S's (i.e., neutral instructions versus instructions to respond overtly). There was, in addition, a group that received no instructions and no review, and another group that was tested prior to seeing the learning material (a training film on map signs). This last was to provide a baseline to ascertain degree of preknowledge of map signs on the part of the S's. Thus, there were eight groups in all. On a test that immediately followed the experimental session, all three variables were shown to have a significant effect, but the expected significant interactions failed to materialize. The significant differences due to the three variables disappeared on a retention test given four weeks later, with the exception of a significant difference due to the sex factor. (See Supplementary Tables 1 and 2.)

Then began a long search for appropriate experimental materials. At that time we did not think of making our own films. Instead, with the hope of making our experiment "realistic" and "practical" we scoured libraries of Army, Air Force and Navy training films. We were looking for a film that contained some manageable material—that is, a content that could be tested for in a fairly straightforward fashion. We finally settled upon what seemed to us to be an excellent film explaining the meaning of twenty conventional map signs (bridges, primary highways, churches, etc.). This film also had a clearly defined review section. We cut this section off and made our own review, since this was one of the variables that we planned to manipulate.

In retrospect it is not quite clear what our motivation was in deciding to run this experiment in the high schools of Nassau County, New York instead of with undergraduates at N. Y. U.; probably it was dictated, at least in part, by our desire to run the experiment on subjects that approximated as closely as possible a group of Army or Air Force trainees. At any rate, this decision involved us in enormous practical problems. Securing the cooperation of high school principals was not difficult—they were very cooperative; but scheduling the experimental sessions was another problem. We were taking intact classes (e.g. a social studies class, an English class, etc.) and running them right in the high schools. The fact that we wanted a retention measure as well as an immediate recall measure required that we have the class on two days exactly four weeks apart. Making out our experimental schedule was easy at first, but it became progressively more and more difficult to fill in the open dates on our calendar. It would turn out that we couldn't get a class at a particular school because the class periods were shortened on that particular day...or, if we could get the class that day, something else prevented the class from being available four weeks later. Our work was further complicated by the fact that once in a while we would lose an entire group because the film broke, or because the sound pick-up didn't work properly, or on one occasion, the class was so unruly that, though we ran the group, we threw the data out without even looking at it.

It seemed doubtful to us at times whether we would be able to finish running the experiment before the schools let out for summer vacation. However, we made it, though without much margin. Once the data were all in, there remained only the task of scoring it. We had tested the subjects by flashing the twenty map signs on the screen one at a time and asking them to write down the name of the topographical feature designated by the sign. This seemed to us to be a fairly straightforward paired-associate task and one that would occasion no difficulties in scoring. This, however, turned out to be a serious misapprehension. The volume of data was so large that we had to hire extra help to score it. Then we found out that scoring these papers was not a routine task. I was frankly astounded at the range of ambiguous responses that a single map sign could elicit. Sometimes the sign for a primary highway would be called simply "highway" or "primary road," or sometimes "p. road." We eventually wound up writing a manual of some thirty pages devoted to criteria for deciding the range of correct responses for each of the twenty map signs. Even then there was a number of cases where the handwriting was so uncertain that the issue had to be decided by consensus.

Simultaneously with this experiment, another experiment was run on "Effect of opportunity and instructions to practice during a training film on initial recall and retention." We used the same films and we also ran the experiment using intact classes from Nassau County high schools.[3] The scoring problems with this experiment were the same as with the other one, but running them simultaneously multiplied the size of the problem of scheduling the experimental sessions.

We learned several things from these two studies that were not mentioned in the experimental reports. One of them, of course, has to do with the scoring problem. We learned that it would be a good idea to select materials in such a way that the responses made by the subjects on tests could be unambiguously classified as either right or wrong. We also learned the folly of bringing the experiment to the subjects rather than the subjects to the experiment. But perhaps the most important piece of general knowledge that we acquired was that face validity—the attempt to make experiments realistic or practical by using actual materials, natural settings, etc., etc.,—is not a good way to get fundamental information, even when the goal is a practical application. The fact that psychological principles may be operating in a real, practical situation does not mean that the real, practical situation is the best one to study if you are interested in discovering the principles. Some of the substantive information that we acquired from the first experiment was utterly trivial (e.g., that people forget over a period of four weeks, that boys are more interested in map signs than girls are, etc.). A borderline case is the finding that increasing the number of reviews raised the immediate recall scores but not

[3]This experiment (reported by Kendler, Kendler, and Cook, 1954) used a 2 x 2 x 4 design, the three variables being: (1) sex; (2) I.Q. (high versus average); and (3) participation conditions. The latter were: (a) overt response; (b) covert response; (c) no instructions to respond (but with maximum time to do so); and (d) no instructions to respond (with minimal time to do so). On a test which immediately followed the experimental session, significant differences were obtained on the sex and participation variables, but not on the I.Q. variable. There were no significant interactions. Of particular interest is the finding of no significant difference between overt and covert participation (cf. Chapter 18), while a significant difference was obtained between instructions to participate (whether overtly or covertly) and no instructions to participate (whether with maximum or minimum opportunity to do so). There was no significant difference between the no-instruction groups which differed on maximum or minimum opportunities to participate. (See Supplementary Table 3 for a comparison of the four participation conditions.) On a retention test administered four weeks subsequently, all three major variables—sex, I.Q., and participation conditions—were found to have a significant effect. Again, there were no significant interactions; and again, within the participation conditions, there was no significant difference between overt responding and covert responding. The only significant differences were between those conditions which allowed for participation (overt, covert, and maximum opportunity groups) and the condition which allowed for minimal opportunity to participate. (See Supplementary Table 4 for a comparison of the four participation conditions.)

necessarily the retention scores four weeks later. Participation, we found out, is quite important for immediate recall, but may be less important for retention, thus suggesting that perhaps it has some motivational value. We found no interaction among the variables of sex, participation, and repetition. It surprised us to find that participation is not increasingly effective as the number of reviews increases. We had thought that it probably would be, but instead we learned that audience participation enhances test performance by a fairly constant amount at all three of the levels of repetition used in this experiment.

In the second experiment (the one on effect of opportunity to practice and instructions to practice), our findings seem somewhat more important in retrospect. We confirmed the findings of several other experimenters that covert (implicit) practice may work as well as overt practice. Again we found a suggestion of the motivating effect of instructions to practice during the film presentation. The overt and covert practice groups scored higher on the immediate recall test than did the maximum-opportunity-to-practice group. But on the retention test four weeks later there was no significant difference between the two instructions-to-practice groups on the one hand, and the maximum-opportunity group on the other. All three of the practice groups (overt, covert, and maximum-opportunity) were superior to the minimum-opportunity-to-practice group on the retention test. One of the things that strikes me is how effective we seem to have been in preventing implicit practice in the minimum-opportunity-to-practice group. The stimulus was on the screen as long for this group as for the others, but in order to keep the subjects from rehearsing it, that is from making implicit responses to it, we filled up the temporal period during which the stimulus was on the screen with non-informative verbiage. Writing the script for this portion of the experiment was a bit of a problem. The verbiage had to sound important, but I could not mention the name of the map sign or describe either it or the feature that it stood for. We thought that this verbal noise would be effective, but, as I recall, we had no clear idea of why. After all, why should listening to something prevent one from talking to one's self? We listen with our ears,—don't we? —not with our vocal apparatus. The answer to the question came from a piece of research far removed from the area of audience participation, and the answer, surprisingly enough, is that we do listen with our vocal cords as well as with our ears.

In the course of investigating synthetic speech sounds, A. M. Liberman (1954) set himself the problem of finding out on what basis we discriminate different speech sounds. Now speech sounds can be classified on at least two different bases. They can be characterized on the basis of their acoustical properties (frequency, intensity, and time) or they can be classified on the basis of the particular set of vocal movements necessary to produce the sounds. Ordinarily, sounds that have similar acoustical properties are produced by much the same sets of muscular movements. But what happens when these two go their separate ways? Which way does discrimination go? Dr. Liberman found that it goes with the muscular movements. The sense that he makes out of his data, therefore, is that his subjects apparently mimic the incoming sounds, and then decide whether they are the same or different on the basis of the kinesthetic feedback from their vocal apparatus. Though Dr. Liberman does not go so far, this finding does suggest that perhaps human beings have an unrecognized tendency to "track" what they hear. That is, they have a tendency to duplicate speech sounds that they hear. If this is the case, then it becomes clear why forcing subjects to listen to uninformative words will tend to prevent them from practicing verbalizations.

I am now willing to believe that human beings tend to verbalize a great many stimuli, not merely speech sounds, but visual stimuli as well, and perhaps other kinds. When S-R interference occurs, that is, when a stimulus prevents a response from occurring, the locus of the interference, I am suggesting, is not sensory but motor, and it occurs because the sensory event has a motor aspect. In some cases this S-R interference does not occur. For example a highly skilled professional typist can sometimes type from manuscript and at the same time answer a few simple questions. ("Is the boss in? No, he'll be back at two o'clock.") In this case, it seems reasonable to me to suppose that the typist is not supplying herself with verbal cues, but rather that she has the manuscript information coded, so to speak, in her hands and fingers. A more subtle application of this interference idea occurs in the interpretation of a later experiment (Cook and Spitzer, 1960)[4] in which it was found that forcing the subjects to copy a visual stimulus with pencil and paper seemed to interfere with their learning that stimulus. This suggests that the implicit response that a subject may use to symbolize a visual stimulus is not necessarily the same thing as the set of cues that are adequate to evoke the response of duplicating that stimulus. For example, perhaps, a particular visual stimulus looks like a gully, and the subject uses the verbal formula "gully" to symbolize the stimulus. But if he has to reproduce that stimulus he may have to say to himself, "Now I start here and draw the line down to there, and then I draw a line up to here." All this theorizing, which may perhaps make some contribution to our understanding of paired-associate learning, serial learning and the programming of teaching machines, owes its origin to these two disparate sources—Liberman's studies in the perception of synthetic speech and the N. Y. U. studies.

In the course of studying audience participation we did an almost incidental little experiment on the effect of prompting upon participation. It wasn't a major effort, but rather a small experiment designed to clear up a minor point in the research. However, it was this little study that changed the direction of the

[4]This experiment employed a 2 x 2 design. The two variables were: (1) prompting versus confirmation, and (2) overt versus non-overt practice. A non-parametric test showed that the non-overt practice condition was significantly superior to the overt practice condition. The conclusion was that overt practice interferes both with learning the response term and with connecting it with its proper stimulus (See Table 2, Cook and Spitzer, 1960).

research.[5] In the review trials following the training film (in the two previously discussed experiments) the subjects were instructed to call out the name of the map sign as soon as it appeared on the screen. To encourage them to do this, the experimenter prompted them on the first three of the map signs by calling out the name himself as soon as the sign flashed on the screen. Then, at Howard Kendler's suggestion, we did a small study comparing the effect of prompting the subjects through all 20 map signs with the effect of no prompting. The difference between the two treatments fell far short of statistical significance, but it started us thinking about the mechanism responsible for the effectiveness of participation. At Kendler's suggestion, I went over all available previous studies in audience participation, and this, I think, included practically all that had been done up to that time. These experiments were studied in an attempt to find out why some experiments showed participation to be effective and others failed to do so.

There were two discernible features that all the previous studies in audience participation had in common, and these were what we later came to call prompting and confirmation. The one that impressed me the more—in fact, almost exclusively—was confirmation. In every case in which audience participation was found to be more effective than no participation the subjects' participation responses were confirmed. That is, information about the correct response was fed back to them immediately after they had made their participation response. Some psychologists speak of this as reinforcement, or knowledge of results. In the case of experiments that failed to find a difference between the participation and non-participation groups, either the participation group had not had the information fed back to them properly or the non-participation group had not been effectively prevented from getting that information.

I had to admit, though somewhat reluctantly and only after a certain amount of prodding from the two Kendlers, that the previous experimental findings were also consistent with the notion that the effective agent in audience participation was prompting. In every case that demonstrated the superiority of audience participation it was possible for the subjects in the participation group to "get a prompt" of one kind or another. For example, in our own experiments using oral (vocal) participation there was nothing to prevent a subject in the participation group from hanging back until he heard what the others were going to say and then chiming in with the voice of the majority. In those studies that failed to find a clear superiority for audience participation it was always possible that the subjects in the non-participation group could have received a prompt.

---

[5]This experiment was concerned with comparing prompting against confirmation. There was no significant difference between the two conditions, although the prompted group performed somewhat better on the more difficult items. The authors concluded that the results suggested that the reason why participation is superior to non-participation is not because the S is given the opportunity to try to anticipate the correct response (as in the confirmation procedure)—which would also allow the S to make wrong responses—but because the S is given the opportunity to practice the correct response. If opportunity to anticipate the correct response was the crucial factor, then prompting should prove to be inferior to the confirmation procedure.

This library research led to the production of two papers, both of which have long since disappeared. One of them listed and classified every method of confirmation that we could possibly think of, and the other one did the same for prompting. Though I could perhaps reconstruct these lists with a little thought, I remember offhand only a few of the methods. For example, "social prompting," or in other words, cheating. Another kind of prompting, as I recall, was "intrinsic prompting." For example, if a subject is assembling a piece of machinery and if he cannot assemble part C until parts A and B have been assembled, then the part A, B, assembly is a prompt for the assembly of C. By the same token the fact that he has correctly assembled C may provide confirmation of the correctness of the AB assembly.

The next experiment that we did (Cook and Kendler, 1956) was designed by Tracy Kendler. It was aimed at testing the relative effectiveness of these two variables—prompting and confirmation at two levels of learning, early and late. (By now we had abandoned all pretense of face validity or practical realism and were going after the important variables.) Recognizing that what we had been working with was paired-associate learning, we decided to continue with this kind of task, but to change the material to try to avoid as many as possible of the problems that we encountered in the audience-participation studies. What we were looking for now was a set of items whose response terms could be reproduced with paper and pencil quickly and easily by the subject but which would be difficult for the subject to verbalize.

By now we had acquired enough sophistication to realize that oral practice responses were evanescent and virtually impossible to control unless we ran the subjects one at a time, which is inefficient. Hence, we planned to use paper and pencil practice, which was proof against inadvertent promptings and confirmations, which provided a permanent record of practice responses, and which enabled us to run our subjects in groups. The insistence upon difficulty of verbalization as a criterion for the selection of experimental materials, at least in the case of my own thinking, represented largely a hangover from the studies in audience participation, where, in order to make sure that the no participation group actually does not participate, it is desirable to make the responses such that subjects would have great difficulty in rehearsing them. The idea that the novelty of the response term was an important variable in paired-associate learning had not yet clearly emerged in our thinking, at least not in my mind. We began by thinking of using just one big complicated figure, but this could not be reproduced quickly or easily by the subject. (This suggestion, which came from Howard Kendler, gave Tracy Kendler the idea of using visual patterns as response terms.) Our problem then became one of inventing visual patterns that met our criteria. The patterns that we finally decided to use were designed in substantially their final form, not by either of the Kendlers or myself, but by the secretary of the project. Each of the response terms, as is shown in Fig. 1, consists of a standard non-representational seven-dot pattern. Each of the first ten letters of the alphabet (the stimulus terms) is represented by connecting three of the dots in the pattern by means of two straight lines. These response terms can be reproduced by the subject in less than three seconds. The format for presenting each item under the prompting condition consisted of the following five events in the order in which they occurred.

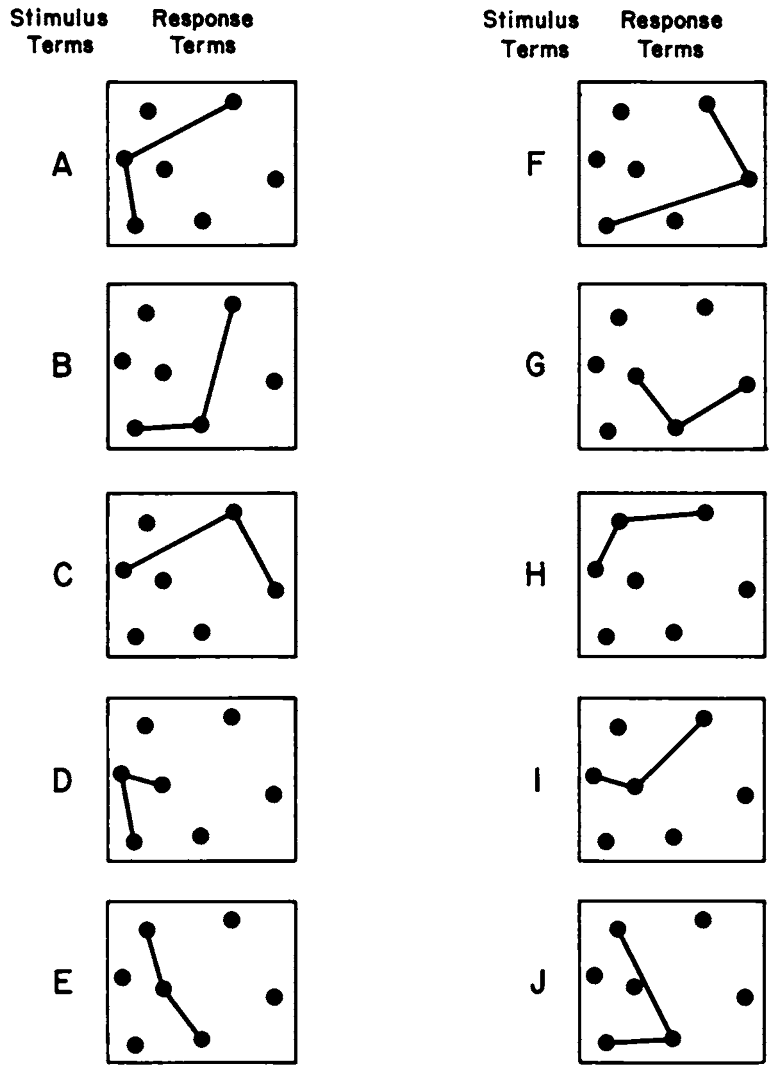


Figure 1. Paired-Associate items used in Cook-Kendler study.

a. The letter (i.e., the stimulus term) appeared on the screen for one second.

b. The screen went black for a period of .25 second.

c. The code line (response term) appeared on the screen for two seconds.

d. The screen went white for a period of three seconds, during which time the subjects drew the response terms on specially prepared sheets of paper containing the dot patterns.

e. The screen went black for a period of one second before the pattern was repeated with the next item.

The total duration of these five events was 7.25 seconds. The events in the confirmation treatment were exactly the same as the events in the prompting treatment. They required a total of 7.25 seconds, and, moreover, each event had the same duration as in the prompting treatment. The only difference was in the order of the events, which was as follows:

a. The letter appeared on the screen for one second.

b. The screen went white for a period of three seconds during which the subjects practiced drawing the code lines.

c. The screen went black for a period of .25 second.

d. The code line appeared on the screen for a period of two seconds.

e. The screen went black for a period of one second before the pattern was repeated with the next item.

The design of the experiment reflected Tracy Kendler's view—based on our analysis of previous experiments in audience participation—that prompting should be effective early in learning when the subject is not spontaneously emitting the correct response, but that confirmation should be better later on in learning because he would then be spontaneously emitting the correct response and hence could profit by practice under exactly the same conditions as those under which he would later be tested. I disagreed flatly. My conviction was that the reinforcing effects of confirmation, whether early in learning or late, would simply overpower all other considerations. The resulting experiment was a 2 x 2 design consisting of the following four treatments: EP (early prompting), EC (early confirmation), LP (late prompting), and LC (late confirmation). The early groups were given a pre-test over the ten items shown in Figure 1, then they were given three prompting (or confirmation) trials with written practice over all ten items. Then they were given a post-test. The late groups were given four "neutral" trials before the pre-test. In the neutral trials the stimulus and response terms for each item appeared on the screen simultaneously for four seconds, and there was a two-second inter-item interval. No practice of any kind was either required or suggested. After the four neutral trials the late groups were given a pre-test which was followed by three prompting (or confirmation) trials and then a post-test. A score for each subject was computed by subtracting the number of items that he got correct on the pre-test from the number of items correct on the post-test, thus isolating the effect of the experimental manipulation.

The results pictured in the bar graph in Figure 2 prove that both Tracy Kendler and I were wrong in our predictions, but I was twice as wrong as she was. Instead of confirmation being better both early and late, prompting turned out to be better. The superiority of prompting was greater for the late group than for the early group, though the interaction was not statistically significant.

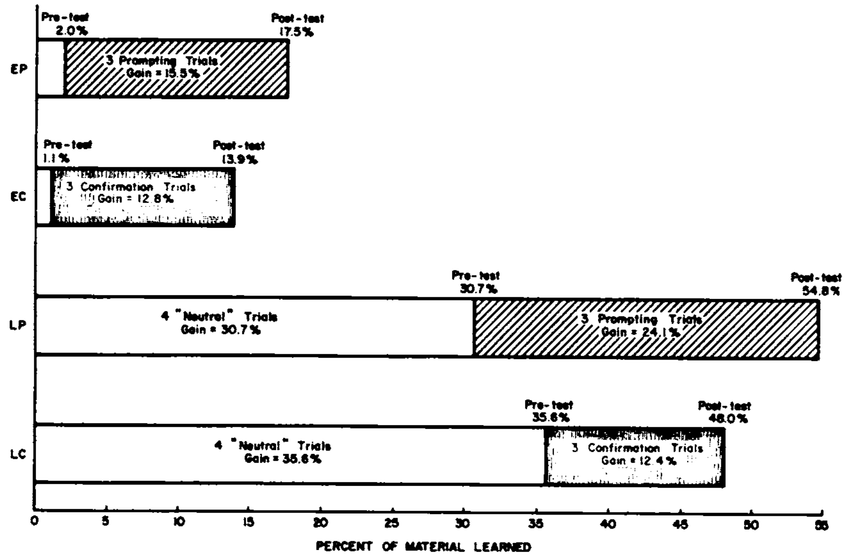


Figure 2. The scores (in percent of material learned) of each of the four cells (Cook and Kendler, 1956)

TABLE 1*

Central Tendencies of the Individual Scores

Cell	N	Mean Correct Responses	Mean Different Legitimate Responses	Median Ratio Substitution Errors to Total Errors	Mean Correct Responses	Mean Different Leg. Resp.
EP	20	.200	1.750	.300	1.550	1.750
EC	18	.111	1.611	.200	1.278	1.333
LP	27	3.074	3.778	.000	2.407	2.222
LC	25	3.560	4.160	.333	1.720	1.760

* From Cook and Kendler, 1956.

We were now face to face with the problem, so familiar to all psychologists, of explaining the unpredicted outcome. It occurred to Tracy Kendler that perhaps the reason confirmation had fared so ill was that the response term had become associated with the stimulus term of the item that followed it during training, so that during the test when a stimulus term was presented it would tend to elicit the response term of the item that preceded it during training. This hypothesis, however, was testable with our data. Since the items were presented in a different random order on each of the three training trials, every stimulus term, except the ones that came first, was preceded by three different response terms. If these three response terms had any tendency to be elicited by stimulus terms that followed them during training, then on the post-test these response terms should occur as substitute responses more frequently than the response terms that had not preceded a particular stimulus term during training. A test of this hypothesis, however, failed to confirm it, and we went back to searching for an explanation.

The theoretical model that we eventually employed as an explanation for our data occurred to me, I remember distinctly, at 20 minutes to five one afternoon in the spring of 1953. I don't know where the idea came from, or why it popped into my head instead of somebody else's. As I recall, about this time I had been thinking about Hull's explanation of problem-solving behavior in rats. Perhaps his emphasis on pure stimulus acts suggested the idea of implicit responses. At any rate it took only two or three minutes to sketch out the theoretical model which is shown in Figure 3. According to this view of paired-associate learning, the subject

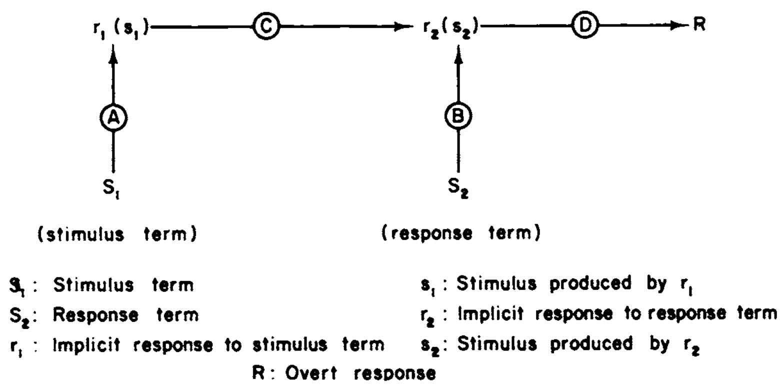


Figure 3. A theoretical model of paired-associate learning.

symbolizes the stimulus term, S_1, to himself in the form of an implicit response, r_1. He does the same with the response term, i.e., symbolizes the response term, S_2, in the form of an implicit response, r_2. The stimulus properties of the implicit response, r_2 (to the response term) cue off the overt response, R_1, of duplicating the response term. In the course of training, this implicit response, r_2 to the response term (S_2) becomes attached to the stimulus properties, S_1, of the implicit response (r_1) to the stimulus term. Thus, there is a mediational chain connecting the two events that occur on the test—the presentation of the stimulus term and the occurrence of the overt response. The model explains the superiority of prompting on the grounds that, in the prompting treatment, connection C between s_1 and r_2 is easier to form than it is in the confirmation treatment because s_1 and r_2 are temporally closer together, and there is no activity intervening between them as there is in the confirmation treatment where the subject is required to make an overt response during this interval.

We very quickly thought of a number of interesting experimental tests of this theoretical model, but it was to be some time before these tests could be carried out, because the contract under which we had been working was terminated, and

I went to work for what was then the Training Aids Research Laboratory of the U. S. Air Force. Being thrown in with a different group of psychologists taught me that the chief obstacle to getting a hearing for the theoretical model was a doubt about the effectiveness of prompting. (A typical response began with a compliment about the ingenuity of the theoretical model and ended with the expression of a doubt that prompting really was a very effective method of teaching paired associates.) In order to allay this doubt before proceeding to further investigation of the intra-item variables in paired-associate learning, it was necessary to perform another experiment. The objection was raised repeatedly that since the best group in the prompting-confirmation experiment had not learned more than about 55 per cent of the material (of Table 1) we were really in no position to say that prompting would be better later in learning. The experiment designed to clear this matter up (Cook, 1958) was a simple two-treatment affair: prompting versus confirmation. Using the materials shown in Figure 1, subjects in the prompting group were given 36 prompting trials with all ten items on each trial and the items in a different random order on every trial. Subjects in the confirmation group were given 36 confirmation trials.[6] Both groups were tested with no knowledge of results after every third trial. The results are graphed in Figure 4. Except for the points at the very beginning and end of the learning curve the differences between the two treatments are statistically significant at high levels of confidence.

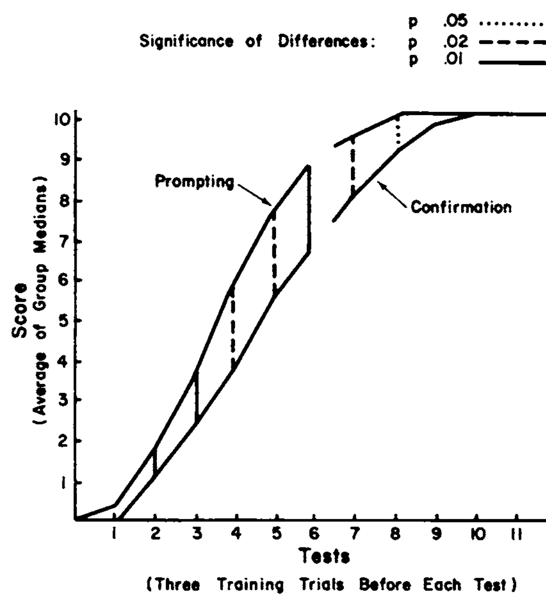


Figure 4. Results of Prompting-Confirmation Study (Cook, 1958)

Thus, the theoretical model was confirmed by the experimental outcome, in the sense that the model proved capable of predicting the results, and the writer knows no other theoretical basis on which the superiority of prompting could have been predicted. The model and/or experimental data related to it have led, heuristically, if not deductively, to some further studies (Cook and Spitzer, 1960; Grier, 1960; Cook, 1960b; Cook, 1961b; Brown, 1960; Angell and Lumsdaine, 1960; Newman, 1960), but it is rapidly becoming apparent that this theoretical model is destined to be supplanted. One of its more obvious inadequacies is the scant attention that it pays to motivation and reinforcement variables. This inadequacy is revealed by the finding of Angell and Lumsdaine (1960) that substituting an unprompted (confirmation) trial for every fourth prompting trial produces results superior to those of straight prompting.

[6] [Special significance attaches to the presence of these test trials in relation to the later study by Angell and Lumsdaine (Chapter 24).]

[See, however, alternative rationale in Chapter 24.] Other inadequacies are revealed and remedied in the more sophisticated model proposed by Newman (1960).

As I look back over this account of our studies, it is obvious that we have come a long way and over a rather tortuous route. It is easily understandable how we arrived where we did, but suppose some sponsoring agency had been interested in the production of a crude model of paired-associate learning. Is it reasonable to think that they could have predicted the route by which we arrived at this destination?

SUPPLEMENTARY TABLE 1

Analysis of Variance on Immediate Post-Test Performance Scores*

Source	SS	df	M.S.	F	P
Sex	165.369	1	165.369	13.764	.0001
Participation	154.908	1	154.908	12.893	.0002
Repetition	113.972	2	56.986	4.743	.019
S x R	29.182	2	14.591	1.214	---
S x P	18.881	1	18.881	1.571	---
R x P	24.614	2	12.307	1.024	---
R x P x S	15.744	2	7.872	---	---
Within Cells	420.538	35	12.015	---	---

*(Kendler, Cook, and Kendler, 1956)

SUPPLEMENTARY TABLE 2

Analysis of Variance of Retention Scores*

Source	SS	df	M.S.	F	P
Sex	198.842	1	198.842	14.889	.001
Participation	53.278	1	53.278	3.989	.054
Repetition	17.705	2	8.852	.663	---
S x R	2.982	2	1.491	---	---
S x P	.928	1	.928	---	---
R x P	18.380	2	9.190	---	---
R x P x S	39.051	2	19.525	1.462	---
Within Cells	467.421	35	13.355	---	---

*(Kendler, Cook, and Kendler, 1956)

SUPPLEMENTARY TABLE 3

The t Values for Differences Between the Means of the Groups Exposed to the Four Different Conditions of Practice (Immediate Recall Test)*

	Overt	Covert	Maximum	Minimum
Overt	-	0.231	2.395**	2.672**
Covert		-	2.598**	2.837**
Maximum			-	0.199
Minimum				-

*(Kendler, Kendler, and Cook, 1954)
**p<.05

SUPPLEMENTARY TABLE 4

The t Values for Differences Between the Means of the Groups Exposed to the Four Different Conditions of Practice (Retention Test)*

	Overt	Covert	Maximum	Minimum
Overt	--	0.047	0.667	3.266**
Covert		--	0.607	3.175**
Maximum			--	2.745***
Minimum				-

*(Kendler, Kendler, and Cook, 1954)
**p<.01
***p<.05

CHAPTER 22

RESPONSE ANALYSIS IN PAIRED-ASSOCIATE LEARNING EXPERIMENTS[1]

John Oliver Cook

The scoring procedure generally used in paired-associate learning experiments throws away most of the information that the experiment actually yields. The experimenter usually counts only the number of items correct (the C score) and perhaps the number of omissions (O score), though occasionally (Gagné, 1950; Gibson, 1942) he may count the number of substitution errors (S score). The need for a more adequate scoring procedure has become urgent now that several experimenters (Cook and Kendler, 1956; Underwood, Runquist, and Schulz, 1959; Murdock, 1959) have recognized that paired-associate learning involves more than a single S-R connection. For example, Murdock (1959, p. 345) says, "...the difficulty with the standard type of paired-associate learning task is that S must learn both the responses and the pairing..." However, as the response analysis to be outlined here will show, Murdock is in error in saying that "...from the data, it is impossible to separate these two different processes." Figure 1 presents a scheme for categorizing all possible "whole" responses in a paired-associate-learning experiment. In some cases S may give a "part" response; for example, if nonsense syllables are employed, and S is instructed to spell the response, he may be able to produce only one or two letters. Such "part" responses as this belong to a sub-class of the appropriate "whole" response category. That is, if S gives one letter and it happens to be the first letter of the correct response, then this letter belongs to a sub-class of the C score category.

In paired-associate-learning experiments stimulus terms and response terms may be selected randomly or in any other way, specified or unspecified, from any class or population whatever. This class itself may be specified or unspecified. For illustrative purposes let us assume that in a hypothetical experiment the stimulus terms are, in order, 1, 2, 3, 4, and 5, and the response terms are the capital

[1] This chapter was prepared by Dr. Cook, now at North Carolina State College, partly as an outgrowth of experiments conducted there and prior experiments conducted at the AFPTRC Training Aids Research Laboratory by him and others. These later studies, as has been indicated by the author, grew more or less directly out of the earlier work reported by Cook in Chapter 21.

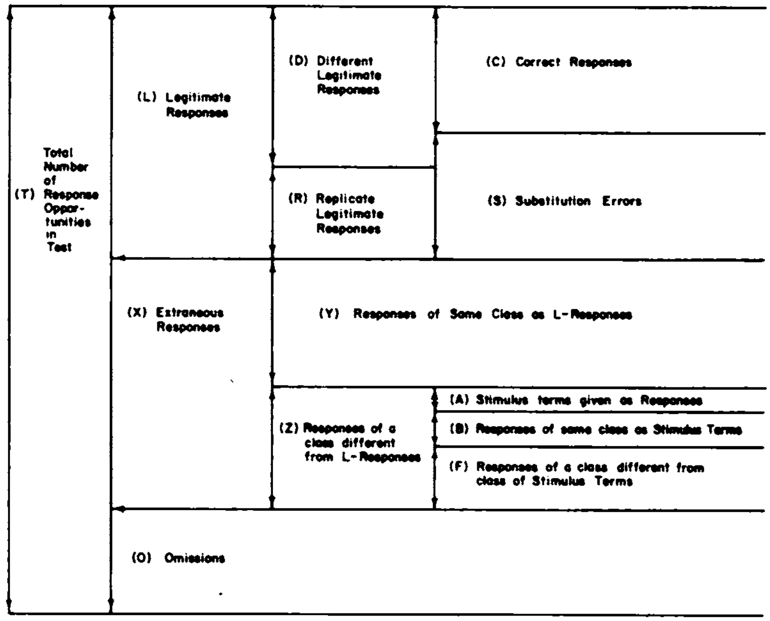


Figure 1. Analysis of responses in paired-associate learning experiments

letters A', B', C', D', and E'.[2] The L score is the number of legitimate responses including replicates. A response is scored as a legitimate response if it is any one of the five response terms in the experiment, whether this response has been elicited by its proper stimulus term or by some other. If the correct responses are, in order, A', B', C', D', and E', and the subject's responses are, in order, A', A', B', D', and E', then his L score is 5. However, only four of these are different legitimate responses. Therefore, his D score is 4. One of the responses is a replicate legitimate response and his R score is, therefore, 1. Since three of the responses, A', D', and E', are correct, his C score is, of course, 3. The number of substitution errors he has made (S score) is L - C, or 2.

Any response other than an omission or one of these five responses is an extraneous response, and the number of such responses is the X score. For example,

[2] It is perhaps worthwhile to state explicitly that the response terms in paired-associate learning experiments are not responses; they are stimuli. And the responses are not response terms. Though I may write as though the subject were emitting response terms, this is not really the case. The responses emitted by the subject may be more or less faithful pictorial duplicates of the response terms, or they may be symbolic equivalents of the response terms. A vocalization of the word "cat" is obviously not a reproduction of the printed word "cat;" rather it is an event that is symbolically equivalent to the printed word according to the equivalence rules of the experimenter's symbol system. We assume, in most cases quite safely, that the subject knows these equivalence rules. However, if he did not know them, this ignorance would be a source of error in his performance.

if the subject's responses, in order, had been A', B', C', P', and Q', his X score would be 2. Since the class from which the response terms have been arbitrarily selected is the class of letters of the Latin alphabet, these two X responses are also Y responses. Had the subject's responses been, in order, A, B, C, 3, 5, the last two responses, which are X responses, would also be scorable as Z responses, since they are of a different class than the L responses. Moreover, since they are also stimulus terms, they would be scored as A responses as well. But had the last two responses been 7 and 8, they would be scored as B responses, since, though they are not stimulus terms, they are terms from the same class as the stimulus terms. However, if the last two responses had been III and V, they would be scored as F responses because they belong to a class that is different from the class of stimulus terms. The nature of this categorizing scheme is such that many scores do not need to be found, since they can be arrived at by computation. For example, S = L - C, B = T - L - O - Y - A - F, Z = X - Y, etc. In some paired-associate experiments, notably those in which stimulus terms or response terms have been selected from unspecified classes, some of these scoring categories will be quite meaningless. In other cases, while it might be possible to compute a certain score, it may be of no theoretical interest to do so.

These response categories (or scores), by retrieving more of the information available in paired-associate-learning experiments, can be useful in a number of ways. Some of the information can be used to test theories; certainly a theory of paired-associate learning that makes any pretense of completeness should predict more than simply the C score. Moreover, since this response analysis provides a scheme for classifying responses that is free from any profound theoretical bias, it can perhaps be useful in clarifying the issues in theoretical disputes, such as the current disagreement about intra-list generalization in paired-associate learning. In addition, when it becomes more widely recognized that two of the basic sources of theories underlying the technology of teaching machines are the areas of serial learning and paired-associate learning, then perhaps a response analysis such as this one, by furnishing a set of error categories, may suggest to the technologist ways in which his programs or displays can be improved.

An example of the way in which these neglected scores—in this case the D score—can be used to throw light upon an experiment is furnished by the Kopstein-Roshal study (1955). The stimulus terms were eight English words and the response terms were their Russian equivalents. A very large number of subjects (N = 2080) was used in comparing two methods of paired-associate training. In what the authors called the "simultaneous" method the stimulus term and the response term appeared simultaneously on the screen for 3-1/2 seconds. In the other method, which the authors called "staggered," the stimulus term appeared on the screen for 3-1/2 seconds. During the last 1-1/2 seconds of this 3-1/2-second exposure, the response term appeared on the screen beside the stimulus term. In both methods, the subjects were given three tests, with three training trials, each containing all eight items, preceding each of the three tests. The simultaneous condition was found to be significantly superior (.01 level) on the first two tests, but not on the third. In explaining the outcome of this experiment it was hypothesized that the simultaneous method, with its 3-1/2-second exposure of the Russian word, provided greater opportunity for the subjects to learn the novel term than did the staggered condition, with its 1-1/2-second exposure of the Russian word. However, if this explanation of the difference in the C scores is correct, it should follow that the C scores should be a direct function of the D scores or, in other words, that the curves for the two sets of D scores should closely parallel their respective curves for the C scores, and that there should be a correspondingly large difference between the two sets of D scores. As Figure 2 shows, this is exactly the case. Hence, we can conclude that this explanation is an adequate one.

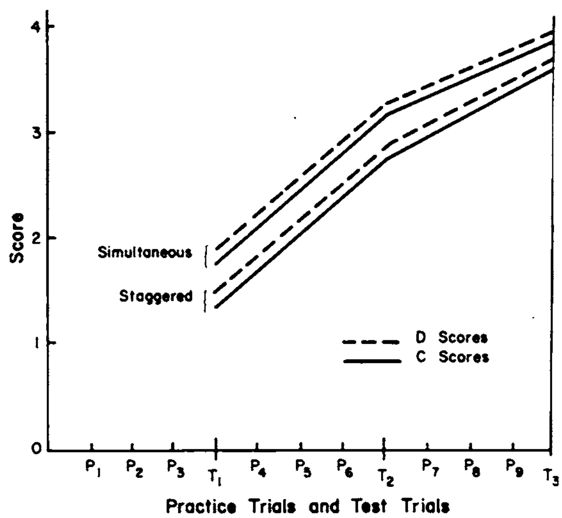


Figure 2. Data from Kopstein-Roshal study of English-Russian word pairs

Another example of the utility of the D score is provided by a recent paired-associate experiment in which the effects of no overt practice of the response term versus overt practice of the response term on two training methods were compared (Cook and Spitzer, 1960). The stimulus terms were the first ten letters of the alphabet and the response terms were non-representational "code lines" that were difficult to learn but easy to reproduce. The prompting method consisted of showing the stimulus term alone for one second, then, after a quarter-second interval, the novel response term alone for two seconds, followed by a three-second period during which the subjects in the overt-practice cell tried to reproduce the response term. Then there was a one-second between-item interval and the pattern was repeated for the second item. In the confirmation method, the duration of events was exactly the same, but the order of events was different. The stimulus term alone was presented for one second. Then, after a quarter-second interval, there was a three-second practice period during which the subjects in the overt-practice cell tried to reproduce the response term. Then the response term alone was shown for two seconds, followed by a one-second between-item interval before the pattern was repeated. As in the Kopstein-Roshal study, all subjects were given three training trials and then a test, with all ten items occurring on each training and test trial, but in a different random order each time. On the test the subjects were shown the stimulus term for two seconds and given four seconds in which to produce the response term. The mean C scores and D scores for the four cells on the test given after the three training trials are shown in Figure 3. Bearing in mind that the D score is an indication of the number of response terms as such that the subjects know, and that the C score is a measure of the number of responses as such that the subjects know and can correctly pair, an examination of these data provide some illuminating insights. For example, the mean C score for Cell I is significantly (.007 level) larger than the mean C score for Cell III. However, the mean D score for Cell III is insignificantly (.24 level)

PRACTICE CONDITION	TRAINING METHOD: Prompting	TRAINING METHOD: Confirmation	Row Means
No Overt Practice	Cell I (N=35) Mean D-Score=3.23 Mean C-Score=2.20 D-C Discrepancy=1.03	Cell II (N=35) Mean D-Score=3.34 Mean C-Score=1.69 D-C Discrepancy=1.65	3.28 1.94 1.34
Overt Practice	Cell III (N=35) Mean D-Score=3.63 Mean C-Score=1.60 D-C Discrepancy=2.03	Cell IV (N=35) Mean D-Score=2.57 Mean C-Score=0.74 D-C Discrepancy=1.83	3.10 1.17 1.93
Column Means: D-Score	3.43	2.96	
C-Score	1.90	1.22	
D-C Discrepancy	1.53	1.74	

Figure 3. Cook-Spitzer data on test given after three paired associate training trials

larger than the mean D score for Cell I. The D-minus-C discrepancy scores for Cell I being 1.03 and for Cell III, 2.03. I think that one can fairly conclude from these data alone that in the prompting method, overt practice of the response term, though it may or may not facilitate the learning of the response term as such, seriously interferes with the pairing of the novel response term and the stimulus term. Looking now at Cell II and Cell IV, it is apparent that with the confirmation method overt practice of the response term does not facilitate the learning of the response term, as such. In fact, it tends to lower both the D scores and the C scores, the difference between the cell means being significant at the .001 level in the case of both scores. In short, in the confirmation method overt practice interferes both with the learning of the novel response terms as such and with the pairing of those terms with their appropriate stimulus terms. Similar comparisons can be made between Cell I and Cell II, where neither the D scores nor the C scores differ significantly, and between Cell III and Cell IV, where both the D scores and the C scores differ significantly (.001 level).

In the case of both the Kopstein-Roshal study and the Cook-Spitzer experiment, a procedure was used that distinguishes between training trials and test trials. However, it goes without saying that the scoring system outlined here is equally applicable to studies that use the anticipation method, in which each training trial is also a test trial.

As suggested earlier, the usefulness of this response analysis in clarifying theoretical issues can be illustrated by the current controversy about "intra-list generalization" in paired-associate learning. This began with a classic paper by

E. J. Gibson (1940). She used the term "generalization error" to designate substitution errors, and, by thus unwittingly smuggling a theory into a label, inaugurated a confusion that has plagued the discussion for the past twenty years. The confusion has recently been compounded by Murdock (1958), who begins by calling substitution errors "incorrect responses." What I call "extraneous responses" he calls "intrusions" but we agree in our use of the terms "omissions" and "correct responses." Translated into the terminology of my response analysis, Murdock asserts the proposition that under proper conditions the number of substitution errors (S score) does not increase and then decrease, but steadily decreases. He dismisses experimental evidence to the contrary on the grounds that the experiments do not meet the proper conditions. The conditions he regards as proper are conditions in which extraneous responses and omissions are not permitted. Under these conditions, he says, if learning occurs, then the number of substitution errors will decline. Murdock's assertion is certainly true, but it is logical truth rather than an empirical hypothesis. Obviously, if the only possible responses are substitution errors and correct responses, and if the number of correct responses increases, then the number of substitution errors must decrease. Murdock's thesis is a tautology.

Actually, there are two genuine scientific issues involved in the controversy, one factual and one theoretical. The factual issue is: What is the shape of the curve of S scores? The contribution that the present response analysis can make to the discussion of this question is to point out that the S score is the numerical difference between the L score and the C score. If the curves of the L and C scores start at zero, then the shape of the S-score curve depends upon the rates at which the other two curves rise. As shown by the data from the four cells of the experiment by Cook and Spitzer (1960) (see Figure 4), these rates, and hence also the S curve, are a

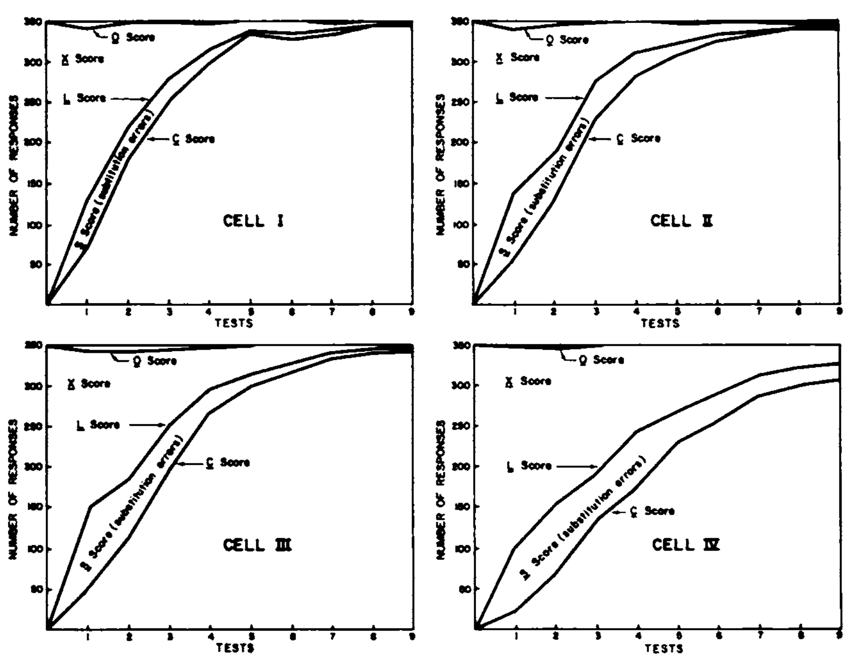


Figure 4. Paired-associate score curves in the Cook-Spitzer study

function of training conditions. They probably are also a function of a number of other variables, but this fact is not brought out by the experiment.

The theoretical issue is that of explaining why substitution errors occur, why the S curve first rises and then falls, and why the substitution errors occur in the particular pattern that they do. Calling them "generalization errors" prejudices the issue but does not resolve it. The question remains whether the phenomena can be adequately accounted for on the basis of stimulus generalization. Certainly generalization theory accounts very handily for the occurrence of substitution errors, and perhaps also for the humped S curve. It is on the third question, concerning the patterning of substitution errors, that generalization theory, at least in its present form, breaks down. A recent paper (Cook, 1958) reported that stimulus terms were not found to elicit each other's responses with anything approaching the equal frequencies that one would expect on the basis that stimulus terms elicit each other's responses because of some sort of similarity. The correlation between the frequencies is actually approximately zero. Instead, the frequency with which a response occurs as a substitute in a substitution error was found to be positively correlated with the frequency with which the response is elicited by its proper stimulus term.

Such findings as this can perhaps be facilitated by using a response analysis that is theoretically neutral, since one is less likely to look into such matters if one is operating under the influence of a strong set induced by a label having theoretical connotations. With minor modifications this response analysis can also be adapted for use in experiments on backward associations where the training is S-R and the testing is R-S. In addition, it can be modified for use in transfer experiments involving two or more paired-associate lists. In both of these cases, it might throw a little light into what are now some very dark corners.

CHAPTER 23

PROMPTING AND CONFIRMATION CONDITIONS FOR THREE LEARNING TASKS EMPLOYING THE SUBJECT-MATTER TRAINER[1]

Leslie J. Briggs

Introduction

Chapter 25 presents the history of the development and experimental use of the device called the Subject-Matter Trainer. References are given there to earlier papers which describe the mechanical details of the device and its various modes of functioning. Much of the early exploratory work dealing with instructions to subjects, pacing of practice, and the role of "position habits" is also briefly summarized there. Chapter 25 also presents a comparison of self-pacing with automatic pacing, which was conducted as a preparation for the studies reported in the present chapter. These are concerned with the use of the various prompting and confirmation conditions, or "modes of operation," provided by the device. (It may be recorded here parenthetically that this study, in turn, was intended to be followed by a study that would adjust both the speed of pacing and the prompting-confirming conditions simultaneously with successive stages of practice and learning. Unfortunately it was not possible to accomplish that study.)

Although experimentation was undertaken to discover the effects of certain single features or modes of operation of the Subject-Matter Trainer (cf. Chapter 25), in practical use of the device the experimentors encouraged students to try out all sorts of combinations and sequences of practice conditions designed to vary total

[1]This chapter is based on a study conducted under Air Force contract No. AF 41(657)-13 with Tulane University under the monitorship of the author, who was then with the AFPTRC Maintenance Laboratory. The chapter is based on an earlier AFPTRC Technical Report (Irion and Briggs, 1957) which was prepared by Briggs on the basis of the manuscript submitted by the contractor. The Subject-Matter Trainer used in this study and in the one reported in Chapter 25 was developed by the Maintenance Laboratory, and the second model of this device, used in the present study, however, was fabricated by the Hathaway Instrument Co. under contract No. 18(1600)-1058. Dr. Briggs is currently director of the Instructional Methods Program at the Northern California Office of the American Institute for Research. Although the present chapter follows chronologically and logically the work described in Chapter 25, it is presented here because of its relevance to the issue of prompting-confirmation dealt with in the preceding chapters (Chapters 21, 22) and the following one (Chapter 24).

practice conditions appropriately as learning progressed. Thus the classroom strategy often made use of the following rules of thumb, conveyed in simple terms to the students:

1. Start the practice in the "prompting" condition (see below) as there is no way to reason out the answers on the first trial (since there are no "stimulus supports").

2. Take plenty of time on the first trial or two in the prompting mode. Review the paired terms covertly several times before going on to the next item.

3. Change to the "confirmation" mode as soon as you feel some confidence in being able to get most of the answers right. Stay in the single-try confirmation mode until you are sure you have most of the items learned.

4. After you gain confidence, and want to have a chance to correct yourself if you do make an error, or if you want to compete with your fellow student to see who has learned best, switch to the multiple-try confirmation mode, and try for the lowest possible time and error scores.

It is apparent that these recommendations to the students were based upon the belief that considerable individual variation may exist in the sequence of conditions that might be best for each student. But in general, the strategy reduced to two points of emphasis: (a) early in practice, go slow and depend on the prompt term; (b) later in practice, try to recall first, then accept the confirmation-correction; speed up the trials as learning progresses, but let the time taken vary in accordance with difficulty in recall for each item.

We shall return, in the discussion, to further considerations for optimizing the practical value of various procedures. But with this brief summary of the rules of thumb employed in classroom use, we turn next to the description of two related experiments having a bearing on testing the validity of these informal practices.

Description of Practice Conditions Employed

Four learning modes and a test mode were employed in the two experiments to be reported here.[2] (Readers not familiar with the Subject-Matter Trainer from other sources—e.g., Briggs, 1958—may wish to refer here to the description given in Chapter 25.)

[2] An apology is made for variations in names applied to the modes of operation in various papers relating to the Subject-Matter Trainer. Some papers attempted to employ terms easily made meaningful to the layman (such as coaching, practice, etc.). In the more technical papers, certain aspects of early terminology from Thorndike's work were followed (such as the "vanishing" and the "retained" situation). Now, as a further change, the terms "prompting" and "confirmation" are adopted as the basic terms to conform with current terminology, with added adjectives and other terms to provide brief yet descriptive labels for the conditions named.

The Prompting Mode

This mode essentially shows the student the right answers at once. When each new stimulus item appears in the window, the student presses a button (called the quiz button). Pressing the quiz button causes a green light located to the left of the correct answer on the response panel to glow. The subject is instructed to associate the correct response with the stimulus item and, following such "covert" responding, he then presses another button (called the stepping button) to bring up the next stimulus. This procedure is repeated throughout the entire series of 20 stimulus items and 20 answers to complete one trial in this mode. Thus the student is instructed not to make an overt response to the stimulus item, but to press the quiz button in order to be shown the right answer. In this mode incorrect overt answers are avoided, but there exists the possibility for incorrect covert responses before the student presses the quiz button or before he looks for the item next to which a green light is glowing.

The Single-try Confirmation Mode

Under this mode the quiz button is not active. Upon the appearance of the stimulus item, the student must press the button next to an answer he chooses. If his answer is correct, the green light next to the chosen item will glow, and the student is then free to press the stepping button to bring up the next stimulus item. However, if the student's response is wrong, a red light glows and a buzzer sounds, indicating that a wrong response has been made. Simultaneously, however, the green light next to the right answer also glows, thus correcting the subject, who is then required to press the right-answer button to acknowledge the correction before the next stimulus item can be presented. In this mode, therefore, it is possible for the student to make only one incorrect overt answer to each stimulus item before the green light shows him the correct answer (which he must then acknowledge).

The Multiple-try Confirmation Mode

In this mode the student may make any number of wrong responses before the correct one is made. After any wrong response the red light flashes and the buzzer sounds, but the green light does not glow to indicate the correct answer to the student. In this mode the correct answer must be found eventually without prompting; then the green light will glow confirming the correctness of the answer, and only then may the next stimulus be presented.

The Single-Feedback Mode

In this mode the student is permitted only one response to each stimulus item, after which he must go on immediately to the next item. After each response, however, he receives either the red-light signal or the green-light signal, indicating whether his response was right or wrong.[3]

The Test Mode

In this mode neither the buzzer nor the red or green lights are activated. The student presses the response button next to his chosen answer, and a counter (not visible to the subject) advances for each wrong answer. In the test mode, the student is given only one try at each answer before being required to go on to the next item, and he is not informed whether or not the response made was correct.

The Three Types of Learning Task

So far, we have discussed the Subject-Matter Trainer only as a device for presenting paired associates.

Although the decision to design the device grew out of observations of the practical difficulty of learning a paired-associate task (component recognition and naming), the response panel layout was determined in view of requirements for serial-learning tasks (memorization of the sequences of steps in brief maintenance procedures). This same design feature (20 response items in view) also is convenient for certain problem-solving tasks. The three types of tasks, or materials, may be learned by the employment of the modes of operation of the device already described.

For paired-associate materials, the stimulus items appear in an arbitrary sequence in the stimulus window, and the 20 correct answers to the 20 stimulus items appear on the answer panel. This same arrangement is used for serial-learning problems, except that in preparation of such problem materials the stimulus items must appear in an order appropriate to the learning task, rather than in an arbitrary order. For problem-solving tasks, an auxiliary display (e.g., map, schematic, or other data) may be placed on a large card in the lid of the device. In this case the questions or problems to be solved appear as usual in the stimulus window, but the student locates his response choice by reference to the material in the lid; opposite his chosen response will appear a code letter which also appears on the response panel. Thus the same response buttons may be used by the student to indicate his chosen answers regardless of which of the three kinds of learning task is being used, and the green lights serve the same function for all learning materials.

[3] During this study the single-feedback mode was placed in effect by using a separate switch to override the student's control of the presentation of stimulus items. After each response is made, the examiner presses the auxiliary switch, bringing up the next stimulus item and thus removing control of item presentation from the subject. For other experimental purposes, this overriding control of the presentation of stimulus items is accomplished by an electrical timer that accomplishes automatic pacing of the presentation of the stimulus items.

Purpose of the Two Studies

The purpose of the first study was to determine the relative effectiveness of the four practice modes of operation of the trainer when each mode was used as the only practice mode throughout a standard training period. This comparison of relative effectiveness was desired for the three kinds of learning materials described above.

The purpose of the second study was to determine whether some combinations of modes of operation, when used in sequence during the practice period, would be more effective than others with respect to efficiency of learning and retention. The procedures and results of these two studies are described next.

The First Study

Since four practice modes of operation of the trainer (prompting, single-try confirmation, multiple-try confirmation, and single feedback) were to be investigated, and since three different problems (paired-associate learning, serial learning, and problem solving) were to be used, a 4 x 3 experimental design was indicated. That is, each mode of operation was used under each of the three problems. A total of 240 subjects was assigned to the 12 combinations of modes and problems (cells) that this design demanded, in order to obtain 20 subjects for each condition.

Subjects. The subjects were male high-school students, all between 16 and 19 years of age. There were two restrictions on the random assignment of subjects to conditions. First, subjects were assigned to conditions so that the frequencies in the cells would be equal, i.e., 20 subjects per condition. Second, the sampling was controlled so that 10 per cent of the subjects in each cell would be Negroes. Subjects were paid for their services.

Materials. The paired-associate problem involved learning the names of 20 map symbols. The symbols (enlarged) were displayed on the response panel, and the names of the things for which they stood appeared in the stimulus window.

In the serial-learning problem, the subject was required to learn the order of 20 steps involved in the processing of photographic film. The 20 steps appeared on the response panel (in unsystematic order, of course). The first stimulus was the question, "What is the first step in developing photographs?" Succeeding questions were phrased, "After you have done ____________, what is the next step?"

In the problem-solving task, a map of a hypothetical region was prepared using the same symbols as were used in the paired-associate task. The numbers 1 through 20 also appeared on the face of the map. The map was displayed above the response panel. The response-panel segments also contained the numbers from 1 through 20. Simple navigational problems were presented in the stimulus window, and the subject was required to learn the correct answer in terms of locating the appropriate numbered symbol on the map.

Procedure. Each subject received 20 minutes of practice on the device. The mode of practice and the nature of the problem, of course, were determined by the condition of the experiment to which he had been assigned. Following this, he was given an immediate test on the device (without knowledge of results). Time and error scores were recorded. Two weeks later, each subject returned and took a delayed test of retention. Again, time and error scores were recorded. Number of errors and test-trial duration (in seconds) were the measures used in the statistical analyses of the test-trial data. Test I and Test II will refer to the immediate recall test and delayed-recall test, respectively.

Results. In general, the error-score means on Test I and Test II show a definite trend among the four modes across learning tasks. A ranking of these error means from least to most goes as follows: prompting, single-try confirmation, multiple-try confirmation, and single feedback. Two exceptions to this pattern occur in the serial and problem-solving tasks on Test II: Single-try confirmation and multiple-try confirmation error-score means are reversed in order from the ranking indicated above.

The pattern of error scores, described in the previous paragraph, was statistically reliable in three instances. A Kruskal-Wallis H test was performed on the modes of each learning task. It was found that the differences among the four modes were significant at the .01 level for the paired-associate task on Test I and Test II, and at the .05 level for the serial task on Test I.[4]

In addition to this over-all statistical analysis of the modes within a learning task, separate comparisons were made between all pairs of modes of those learning-task groups that had been found to be reliably different. In Tables 1, 2, and 3 are presented the probability values[5] (p-values) of these mode pairs.

Table 1, referring to the error-score data for the paired-associate task on Test I, shows that (a) the prompting mode is significantly superior (i.e., fewer errors were made) to the other modes, and (b) the single-try confirmation versus single-feedback comparison is significant. The same results also appear for Test II data (on the same learning task), except for the difference between the single-try confirmation and prompting modes, which is at a marginal level of significance. (See Table 2.)

[4]Not included in the results presented here, but a feature of the experimental procedure, was the administration of the Otis intelligence test to subjects in the first experiment. But the Otis scores were not used to alter the random assignment of subjects to conditions. Nor were Otis scores used in an analysis of co-variance, as originally planned, due to (a) the lack of a relationship between Otis scores and experimental test scores and to (b) skewness and heterogeneity of variance of the transformed test-trial scores.

[5]These p-values were obtained with a rank test (Edwards, 1954) that tests for the significance between two groups of scores; it is a special case of the H test.

TABLE 1

P-Values (Two-tailed) for Pairs of Modes of the Paired-Associate Learning Task, Using Number of Errors Made on Test I

Mode	Single-try Confirmation	Multiple-try Confirmation	Single-feedback
Prompting	.035	.001	<.001
Single-try confirmation		.097	.013
Multiple-try confirmation			.429

TABLE 2

P-Values (Two-tailed) for Pairs of Modes of the Paired-Associates Learning Task, Using Number of Errors Made on Test II

Mode	Single-try Confirmation	Multiple-try Confirmation	Single-feedback
Prompting	.094	.027	<.001
Single-try confirmation		.349	.014
Multiple-try confirmation			.439

TABLE 3

P-Values (Two-tailed) for Pairs of Modes of the Serial-Learning Task, Using Number of Errors Made on Test I

Mode	Single-try Confirmation	Multiple-try Confirmation	Single-feedback
Prompting	.308	.132	.001
Single-try confirmation		.882	.076
Multiple-try confirmation			.049

The p-values given in Table 3 for the serial task (Test I) show a different picture. The single-feedback mode is significantly inferior (i.e., most errors were made) to the other modes—with the exception of the single-try confirmation versus single-feedback comparison which, however, approaches significance; but the other mode pairs are not reliably different. This result, together with data in Tables 1 and 2 on the paired-associate task, suggests that possibly the difficulty of learning the task is a factor in the relative training efficacy of the various modes, apart from the type of learning task. The raw data (not shown here) do indicate that, in terms of number of errors, the serial task was more difficult than the paired-associate task.

Of secondary interest are results in terms of test-trial duration. Although the data are not presented in this report, test-trial duration means and medians on Test I and Test II demonstrate two different patterns among the four modes within a learning task. One pattern (Test I) is similar to that of the error-score means, which was described earlier; the four modes of operation can be ranked from shortest to longest in associated test-trial duration (in seconds) as follows: prompting single-try confirmation, multiple-try confirmation, and single feedback. This variation among correlations was statistically reliable at the .05 level (using the Kruskal-Wallis H test) for only the paired-associate task on Test I. The other pattern (Test II) is that test-trials for multiple-try confirmation mode have the longest duration (i.e., largest mean and median), but that the results for the remaining modes are somewhat similar, although there is a tendency for the prompting mode to have the shortest duration. This pattern was statistically reliable at the .05 level only for the serial task on Test I. No further breakdown of the duration data was attempted.

The Second Study

The second study was concerned with trying to determine whether or not particular combinations of modes of practice produce greater efficiency of learning and retention. One unfortunate circumstance exists with respect to the second study. It was not possible to choose the type and number of practice conditions desired for the second study until the results of the first study were available. It was therefore not possible, at the outset, to draw subjects from a single population for random assignment to the various conditions of both studies. This fact, in turn, makes it impossible to determine whether or not a single mode of presentation is superior to a particular combination of modes of operation.[6] Of course, subjects were assigned to conditions at random within studies (with the single restriction that there would be a fixed percentage of Negro subjects in the sample).

[6]A rough comparison might nevertheless have been made were it not for the fact that subjects for the second study were selected from a somewhat different population than the subjects for the first. This was because, among other things, most of the work on the second study was conducted during the summer. Few high-school students attend summer school unless they have failed one or more subjects during the regular school year. As a result of this fact, subjects in the second study were probably inferior to subjects in the first study—a fact that would invalidate the above-mentioned comparison.

Subjects. The subjects were all males, aged 16-19 years. They were recruited from high schools, both public and parochial (including one municipal trade school), and from a variety of boys' organizations. Subjects were randomly assigned to the nine cells (conditions) of the experimental design, with the restriction that each cell contain 18 white and two Negro subjects.

The nine cells represent combinations of three learning tasks with three practice procedures. The three learning tasks were identical to those used in the first study, i.e., a paired-associate task, a serial-learning task, and a problem-solving task.

Materials. The same learning tasks were used in the second study as had been used in the first. The procedure for giving different combinations of modes of operation of the trainer follows.

Procedure. The three practice procedures were as follows:

1. Prompting followed by single-try confirmation. Subjects were permitted to practice for 10 minutes using the prompting mode, followed by 10 minutes on single-try confirmation, with no break in time between the two modes except for the few seconds required for the experimenter to inform subjects that it was time to change procedures.

2. Prompting followed by multiple-try confirmation. Subjects practiced 10 minutes on prompting, followed by 10 minutes on multiple-try confirmation mode.

3. Prompting followed by single-try confirmation and then by multiple-try confirmation. Subjects practiced five minutes on prompting, five minutes on single-try confirmation, and 10 minutes on multiple-try confirmation.

Under all conditions, subjects were given preliminary instructions describing the nature of the learning task, the practice procedures, the length of time they were to practice on each, and the instructional phrases that would signal changes in procedure.

Each subject practiced for a total of 20 minutes, followed immediately by a test on the device. No knowledge of results was given during the test. Time and error measures were taken. Two weeks later, subjects were retested with the same materials used on the first test. Time and error measures were again recorded.

The obtained time and error measures were tested by approximate statistical procedures to determine whether the different practice combinations produced different degrees of learning.

Results. Inspection of means for time and error scores for Test I (immediate recall) and Test II (delayed recall) revealed differences among the various learning tasks. This result is not important, as the three tasks were different not only as to type of task, but also as to subject matter. No effort had been made to equate the three tasks for difficulty.

This study sought to determine whether different combinations of practice modes result in different degrees of effectiveness in training. Separate analyses of the results were made within each of the three learning tasks for Tests I and II. None of the six tests revealed any statistical evidence of differential effectiveness among the three practice-mode combinations.

Inspection of the means and medians of the arrays of scores revealed no consistent pattern nor any marked degree of skewness, so no transformation was made. The scores within each learning task were tested for homogeneity of variance. For the time scores, it was found that the variance was homogeneous in only three of the six cases. For these scores, therefore, a non-parametric rank test (Kruskal-Wallis H test) was used to test for differences.

For the error scores, the variances were homogeneous in all cases. These scores were treated by analysis of variance to test for differences among practice-mode combinations.

Neither error nor time scores, when treated as indicated above, revealed any significant differences among conditions, although for both types of score, the group that had prompting followed by single-try confirmation and then by multiple-try confirmation showed slight superiority over the others.

Discussion

The results of the first study have indicated that, in general, the four modes of operation tested with the Subject-Matter Trainer are arranged in descending order of effectiveness as follows: prompting, single-try confirmation, multiple-try confirmation, single feedback. The degree of differences among these modes and the significance of those differences, however, varied according to the three types of learning task employed.

Perhaps the most significant result of the first study was the general superiority of the prompting mode of operation, which requires no overt response and may appear to be a more passive mode of rehearsal than the others. The superiority of this mode may be accounted for by a combination of four factors:

1. No time is lost searching for the correct response term, and no feedback time is involved. Thus more practice trials are possible in a given practice period. Since the lesson seems to go along easily, a subject may take time on each item to "rehearse covertly" and to supply for himself stimulus supports of his own construction.

2. Overt wrong responses are avoided, although covert wrong responses are possible if the subject fails to follow directions precisely.

3. No signal (the buzzer and the red light) for a wrong response is ever received, since no overt responses are made.

4. There is close temporal contiguity between presentation of the stimulus and the appearance of the green light.

Regardless of differences among the three learning tasks in relative effectiveness of the four modes of operation, what would seem to be an effective level of learning did take place in all modes for all learning materials. No direct comparison is available between rate of learning on the device and that achieved by more conventional means, but it is interesting to note that many of the mean errors on test trials ranged from only .5 to about 5.0, and very few means exceeded 10.0. Without regard to the differences in effectiveness which appeared among the four trainer modes of operation, a possible implication of the first study is that any systematic means of rehearsal of the materials with either prompting or confirmation is an effective way of learning. Thus "active (including covert) responding" may be just as characteristic of the prompting mode as of other modes. It is interesting to note that this mode requires neither "constructed" responses nor "multiple-choice" responses.

The second study, concerned with testing the relative effectiveness of three combinations of modes of operation when used in succession for the three types of learning task, revealed no differences among the nine conditions thus established. At first glance this may appear surprising, considering that some significant differences were found in the effectiveness of the four single modes of operation. It is believed that the explanation for the lack of differences among combinations lies in the fact that too much practice time was allowed in the prompting mode under each condition before other modes were introduced. That is, since the prompting mode is an effective mode of learning, most of the mastery of the material was probably achieved before the second and third modes in the sequence were added. In retrospect, it appears to have been a mistake for the purpose of this experiment to allow the 10 minutes in the prompting mode that was employed under two conditions for each task, and the five minutes in the prompting mode that was employed under the third condition. To test more effectively for differences among mode combinations, therefore, the prompting mode, if employed first in each case, would have to be reduced to a minimum. In view of the results of the first study, and in view of our concept of the use of the prompting mode, we thought it not reasonable to employ the prompting mode last rather than first in the sequences of conditions. Doing so could well have resulted in some significant differences among combinations.

On the other hand, for practical purposes, the superiority of prompting, whether alone or in combination with other modes, appears the most important finding of the studies. It is suspected that mode combinations, to be superior to prompting alone, may have to be "tailor-made" for the material, or must involve much more experimentation to develop or find combinations superior to prompting alone.

The results of the study show only slight and non-significant evidence that changing from prompting to other modes as trials progress (e.g., to single-try confirmation and then to multiple-try confirmation) may be preferable to continued practice in the prompting mode of operation. However, such a sequence clearly would be preferable to practice only in the non-prompting modes (first study). Thus the total results are not inconsistent with the belief that practice should begin with prompting followed by confirmation, the procedure the experimenters used often in the classroom (Briggs and Besnard, 1956), though they do not offer strong or direct support for this view. The clearest purport of the present findings is the support they furnish for the findings of Cook (1958: see Chapter 22) as to the effectiveness of prompting used alone. The later finding of Angell and Lumsdaine (1960: see Chapter 24) concerning alternation of prompting and confirmation can, of course, be interpreted in the light of the mixed-modes rationale that has been presented here.

For best results, even more flexible sequencing of conditions should be possible. That is, whereas Angell and Lumsdaine (1960) introduced confirmation to replace prompting in every fourth trial, it would appear that optimum individual sequences could be worked out to fit the subject and the material more precisely. In so doing, the results of experimentation should be a helpful guide.

It is to be noted that the serial task used in this study was more difficult than the paired-associate task. Due to its nature, this task presents a problem that is not most effectively dealt with by whole-trial practice. Subsequent to the experiments reported here, a modification was made in the device that would permit the subject to practice short segments of the task at first, and, as practice continues, to select any successive parts of the task in any size unit desired (cf. similar procedures studied by Maccoby and Sheffield—e.g., Chapter 5). This was accomplished by placing 20 switches in a series so that the subject could press any switch to present the corresponding step as a starting point for part-task practice. This modification presumably could be employed to study the "memory-span" of units most effective for learning, and to permit individual subjects to practice the sub-sequences most in need of improvement.

A second device modification, useful for paired-associate but not serial learning, was a series of memory circuits that caused stimulus items to be "skipped" or dropped out of the series when they had been responded to correctly on two consecutive trials. In fact, by setting a special switch, any desired number of successive correct responses could be selected as the criterion for "item drop-out."[7] A similar feature was incorporated into the so-called Card-Sort Device (Briggs, 1958), except that there the correct items are not only dropped out, but the remaining items can be "shuffled" to break the sequence, or left in a fixed sequence, as desired. In both devices, the length of list thus reduced toward zero as items are learned. This economical feature should make a great difference when time to mastery is the

[7][A similar feature was provided by mechanical means in devices described by Pressey (1927) and Skinner (1958).]

criterion. On the other hand, retaining all items for all trials, as in the present experiments, probably leads to overlearning of some items as remaining items become learned.

Conclusions

1. When employed as the sole practice condition, prompting is superior to various confirmation conditions, the amount of superiority varying with the task, the specific confirming conditions employed, and, probably, with stage of mastery.

2. When prompting is employed first, followed by various confirming conditions, there appears little difference in effectiveness, whether the rest of the sequence employs (a) single-try confirmation, (b) multiple-try confirmation, or (c) single-try followed by multiple-try confirmation. Some slight but non-significant favor is indicated for sequence (c), above.

3. No sequence was studied in which confirmation conditions were followed by prompting, since there appeared to be logical reasons to believe it would be a poor procedure because no prior study of materials preceded the practice. In such an event, the optimum sequence presumably would depend upon the degree of learning resulting from the "study."

4. Realization of the expected superiority (over straight prompting) of a correlation in which early prompting is followed later by switching to confirmation might require tailoring the procedure to specific materials and subjects.

5. Two modifications of the Subject-Matter Trainer, which should improve its effectiveness for paired-associate and for serial learning tasks, are described. These pertain to "item drop-out" for learned items in a paired-associate task and to whole-part practice sequences for serial tasks.

CHAPTER 24

PROMPTED AND UNPROMPTED TRIALS VERSUS PROMPTED TRIALS ONLY IN PAIRED-ASSOCIATE LEARNING[1]

David Angell
A. A. Lumsdaine

Cook and Kendler (1956) and Cook (1958)—see Chapters 21 and 22—have reported results of experiments that show the superiority of a "prompting" procedure over a "confirmation" procedure for learning paired associates. In the former (prompting) procedure, the learner sees the stimulus term, then sees the response term, then copies the response term. In the latter (confirmation) procedure, the learner sees the stimulus term, attempts to produce the response term, and then sees the response term. The prompting procedure thus resembles a classical conditioning trial, while the confirmation procedure resembles an operant conditioning trial and is similar to the standard "anticipation method" of learning paired associates that has been widely used in the psychological laboratory. Cook has shown that the prompting procedure is the more effective of the two, not only in early learning trials (Chapter 21) but also in later trials (Chapter 22). Further, its effectiveness evidently does not even depend upon the learner's making an overt response (Cook and Spitzer, 1960; cf. also the preceding chapter by Briggs).

The results of Cook and his associates are important because they bear directly on a fundamental question in the technology of teaching and training, and because they have major implications for the construction of auto-instructional, or teaching-machine, programs. A basic aim of all teaching is to get the learner to make responses to stimuli which, prior to the educational or training process, would not evoke those responses. An obvious first requirement for training is to get the learner to make the desired response in the presence of the to-be-conditioned stimulus. Ordinarily, this is done by presenting the to-be-conditioned

[1]This chapter is reproduced with minor editorial changes from a Technical Note (Angell and Lumsdaine, 1960b) reporting the first of a series of experimental investigations of cuing variables in progress at the American Institute for Research under the sponsorship of the Air Force Office of Scientific Research, under contract No. AF 49(638)-681. A preliminary report on this study was presented at the first annual meeting of the Psychonomic Society in Chicago in September, 1960 (Angell and Lumsdaine, 1960a). The authors are grateful to John O. Cook for making his experimental materials available and for explaining his apparatus design to the authors. The assistance of Richard A. Hyman, who saw to the construction of our materials and helped run the subjects, is also acknowledged with gratitude.

stimulus in temporal and/or spatial contiguity with a stimulus that has the power (usually by virtue of prior learning) to evoke the desired response. By having the response and the to-be-conditioned stimulus occur together, the connection between them can be strengthened, so that the probability of the response being evoked by the stimulus will increase from zero, or close to zero, at the beginning of training, to close to 1.0 at the end of training.

However, since initially the response is elicited by cues (stimuli) that will not be present in the criterion situation, questions concerning the amount and kind of stimulus support provided during training are of crucial importance. These may be of many types, as shown, for example, on the detailed analysis by Meyer (1960) of cues used in one of her programs. It seems reasonable to suppose that some cuing methods (i.e., techniques of utilizing the cue-support feature of training) will lead faster and more reliably to error-free responding that is independent of the supporting prompts than will other methods. Accordingly, the general aim of the research program of which the experiment reported here is a part is to determine optimal cuing methods for achieving this outcome with self-instructional programs.

Skinner, a pioneer in the development of automated teaching methods and programs, recommends a procedure of progressively, and gradually, diminishing the amount of stimulus support provided on successive presentations of the material (1958). The technique, which he calls "vanishing," is illustrated in the teaching-machine programs constructed by Skinner and his associates, and by numerous other program writers following his general approach. The elementary-level spelling program by Porter (1959) provides a good example; there are by now numerous others. This gradual "weaning" of the learner from stimulus support that is initially essential but terminally undesirable, is, of course, no innovation in teaching (cf. Briggs, 1960). For years, teachers and tutors have used hints and prompts that are strong and direct at first, and subtle and weak later. What _is_ new in the Skinner programming method is the opportunity for fairly precise manipulation of cuing factors[2] and the resulting possibility of empirically determining optimal cuing arrangements.

However, there appears to be an implication from the results of Cook's several studies that efforts to get a learner to respond in the presence of less than total stimulus support may be unnecessary. In his experiments, those subjects learned fastest who did not attempt to produce a response with less than a 100 per

[2]The prompts used in Skinner-type teaching programs are contextual and semantic in nature. Grammar, syntax, and vocabulary are used in ways that constrict the population of appropriate responses to one or two fairly obvious possibilities. (Cf. Skinner, 1958; Skinner and Holland, 1958; Meyer, 1960.) The research program under which the present study was conducted is also investigating forms of prompting that are achieved by manipulation of physical cue properties such as clarity and duration.

cent prompt available. The issue is thus raised of whether it may be more effective to have the learner simply copy an answer a number of times, rather than to have him attempt an answer on the basis of the partial cues provided, and then to have that constructed answer confirmed. Cook's results do not clearly settle this issue, however. For one thing, there remains the possibility, as yet untested, that some combination of prompted and unprompted trials is superior to a pure prompted sequence. Secondly, the experimental procedure used by Cook did not provide a test of a "pure" prompting condition, since every fourth trial throughout the learning sequence was a "test" trial, in which the learner attempted to produce the response term without its being presented. Twenty-five per cent of the trials in Cook's prompting condition were, thus, unprompted trials.

The present experiment was designed to yield data to test the hypothesis that a "pure," or, more exactly, a "purer" prompting sequence will be less effective than a "mixed" sequence in which both prompting and confirmation-type trials occur. The basis for this expectation is that the mixed sequence not only insures practice of the correct response, but also provides practice in the desired terminal criterion condition of responding correctly in the absence of prompting. Essentially, this experiment consists of a replication of Cook's experiment (see Chapter 22), with procedural modifications that provide for comparing the effects of a purer prompting condition with those of a condition that combines prompting and confirmation trials.

Method

Materials

Stimulus materials were the same as those of Cook. As illustrated in Figure 1, the stimulus term of each of the ten S-R pairs was one of the first ten letters of the alphabet; the response term was a "code-line" pattern consisting of two straight lines that connected three of the dots in a standard field of seven randomly placed dots. The material was projected on a viewing screen by a 35-mm. automatic film-strip projector. The timing of intra-trial and inter-trial events was controlled by magnetic tape running through a tape recorder at 3-3/4 inches per second. The tape was programmed with aluminum foil patches that closed an electrical contact to activate the projector's frame-changing mechanism. The trials were of three types—two types of practice trials, and a test trial. These are diagrammed in Figure 2. Subjects responded by drawing the code-lines to connect three of the seven standard dots in boxes appearing on mimeographed answer sheets, which were stapled together into booklets of twelve sheets.

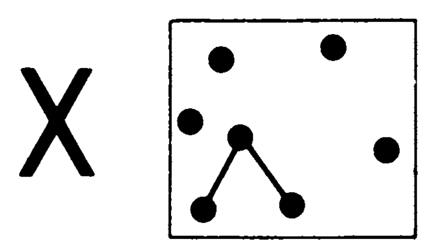


Figure 1. Illustrative paired-associate item, pairing a letter ("X") as stimulus term with a code-line pattern as response term.

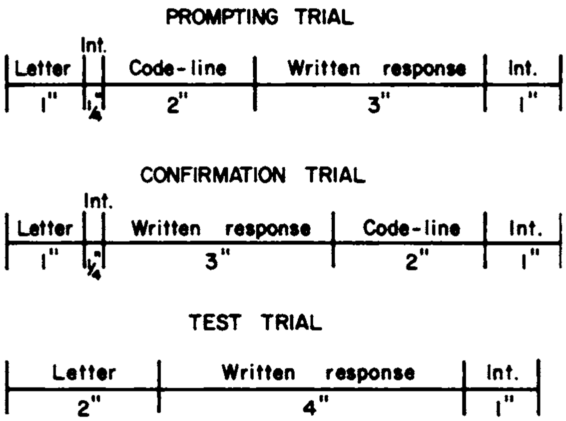


Figure 2. Sequence and duration of events during presentation of one paired-associate item for each of the three types of trials.

Subjects and Procedures

Data will be presented for six groups of subjects (S's)—three groups in each of two experimental conditions. Most of the subjects were high school students, several were first-year college students, and several were non-students. Median age of S's in the "prompting-alone" condition was 17-10, range 15-11 to 25-3, and of S's in the "prompting-plus-confirmation" condition, 17-5, range 16-0 to 20-10. All subjects were paid volunteers. Subjects were randomly assigned to the experimental conditions. N's for the three prompting-alone groups were 8, 11, and 11, for a total of 30; N's for the three prompting-plus-confirmation groups were 10, 5, and 11, for a total of 26.

Each group received 24 practice trials and three test trials. A trial consisted of ten S-R-pair presentations. The first test followed the 12th practice trial; the second test followed the 24th practice trial; and the third test was administered 15 minutes after Test 2. Random activity intervened between Tests 2 and 3.

Experimental Conditions

In the prompting-alone (P) condition, all 24 practice trials were of the prompting type, that is, S's saw the letter, then the code-line pattern, and then drew the code-line. In the prompting-plus-confirmation (P+C) condition, three-fourths of the trials were of the prompting type and one-fourth of the unprompted ("confirmation") type. On every fourth trial in the P+C condition, S's tried to produce the code-line pattern corresponding to a particular letter, before seeing the code-line pattern. In the test trials, only the letter was shown and S's drew the code-line patterns without seeing them on the screen. Test-trial responses were recorded on separate sheets, the booklets of practice sheets having been previously collected. The experimental sessions were just under two hours in length.

Figure 3 shows a comparison of the prompting-alone and prompting-plus-confirmation conditions in the present experiment; and between the prompting-and-confirmation conditions in Cook's (1958) experiment. Each letter represents a trial consisting of the presentation of 10 S-R pairs. The letter indicates the type of trial, where P = prompting, C = confirmation, and T = test (cf. Figure 2).

Present Experiment

Condition	
Prompting Alone	PPPP PPPP PPPP T PPPP PPPP PPPP T T
Prompting-Plus-Confirmation	PPPC PPPC PPPC T PPPC PPPC PPPC T T

Cook's Experiment

Condition	
Prompting	PPPT PPPT PPPT PPPT PPPT PPPT PPPT PPPT PPPT PPPT PPPT PPPT
Confirmation	CCCT CCCT CCCT CCCT CCCT CCCT CCCT CCCT CCCT CCCT CCCT CCCT

Figure 3. Experimental procedures of the present experiment and of Cook's (1958) experiment compared.

Results

Table 1 shows the number and per cent of persons in each group, in each condition, that achieved perfect scores on each test.

TABLE 1

Number (#) and Percent (%) of Persons with Zero Errors on Each Test

Prompting Alone							
		Test					
	Total	1		2		3	
Group	N	#	%	#	%	#	%
P-1	8	1	13	4	50	2	25
P-2	11	1	9	5	45	4	36
P-3	11	3	27	7	64	8	73
Totals	30	5	16.7	16	53.3	14	46.7

Prompting-Plus-Confirmation							
		Test					
	Total	1		2		3	
Group	N	#	%	#	%	#	%
P+C-1	10	3	30	6	60	5	50
P+C-2	5	1	20	4	80	4	80
P+C-3	11	3	27	8	73	9	82
Totals	26	7	26.9	18	69.3	18	69.3

Chi squares were calculated to test the hypothesis that the three groups in each of the two experimental conditions were random samples from common populations. Since this hypothesis is not refuted for any of the tests, the data for the three groups in each condition are pooled and treated as a single sample for further analyses.

In Figure 4, the pooled data for the three groups in each of the two conditions are plotted. The percentages are those shown in the bottom row of Table 1.

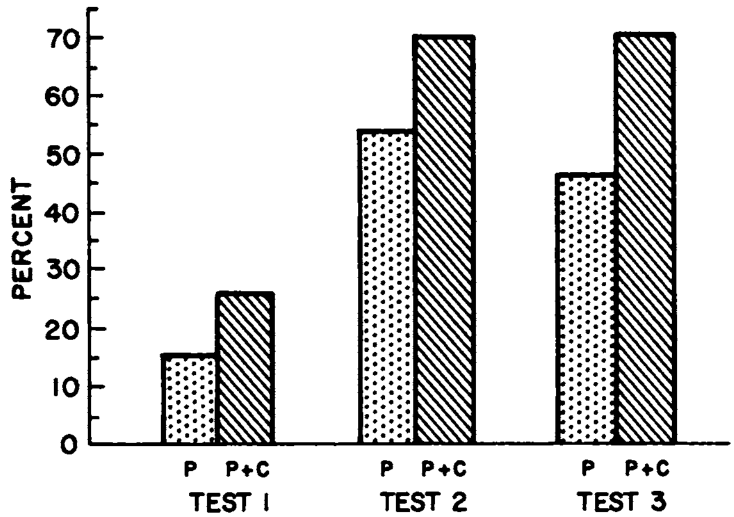


Figure 4. Per cent of persons with zero Errors.

Chi squares were calculated for each test, the data being numbers of S's in each of the two conditions that achieved perfect scores. Only on Test 3 is there an intimation that the difference may be due to something other than random sampling: the χ^2 for Test 3 yields a P-value of <.10.

Note, however, that the absolute values of the several differences are not small, and that they are all in the same direction. On each test, a higher proportion of people in the P+C condition showed errorless performance.

The differences, it can be seen, are larger on the two terminal tests than on the early test. The amount of improvement in error-free responding from early in practice to late in practice appears to be greater in the P+C condition than in the P-alone condition. To check this further, the results on Tests 2 and 3 were examined for those S's who made one or more errors on Test 1. The data in Table 2 show the proportions of persons who made errors on Test 1, but who made no errors on the later tests.

TABLE 2

Proportions, Differences Between Proportions, Z and P Values for S's Making One or More Errors on First Test, but No Errors on Second and/or Third Tests

No errors on Test Number:	Condition P	Condition P+C	Diff.	σ_D	Z	P
2	.44	.58	.14	.152	0.92	.36
3	.36	.58	.22	.152	1.45	.14
2 or 3	.28	.53	.25	.148	1.69	.09

The differences, again, are not small. The largest of them achieves a level of significance sufficient to permit a moderate degree of confidence for rejection of the null hypothesis (which here would state that there is no real difference between proportions of S's in the two conditions who, having made errors on the first test, then showed error-free performance on the later tests).

At the end of the practice trials, a large enough number of S's in both conditions had mastered the paired-associate task so that the error distributions on the two terminal tests are badly skewed. Considered by themselves, these data are not readily amenable to statistical interpretation.

The error distributions for Test 1, which occurred midway through the practice trials, are more nearly normal. Figure 5 is a histogram showing the error distributions for S's in the two experimental conditions.

For the error data of Test 1, the Mann-Whitney U Test was used to test the hypothesis that the two error samples have the same distribution. The value of z, calculated by the Mann-Whitney method, is 1.67. The two-tailed probability of this z score is .095.

When the total number of errors made by all S's in both conditions on all tests is considered, the median number of errors is one. In the P-alone condition, there are 35 cases of less than one error; in the P+C condition there are 55 cases. For these figures, the χ^2 for the Median Test is 3.85, which, with df = 1, is just significant at P = .05. Over all three tests, then, there are significantly more cases of perfect scores occurring in the P+C condition than in the P-alone condition.

The probabilities cited have been two-tailed. Since our basic hypothesis is unidirectional, viz., that S's run in the prompting-plus-confirmation condition will show superior performance to S's run in the prompting-alone condition, and since any differences in the opposite direction would constitute refutation of our hypothesis, it is not indefensible to use one-tailed probabilities to evaluate the

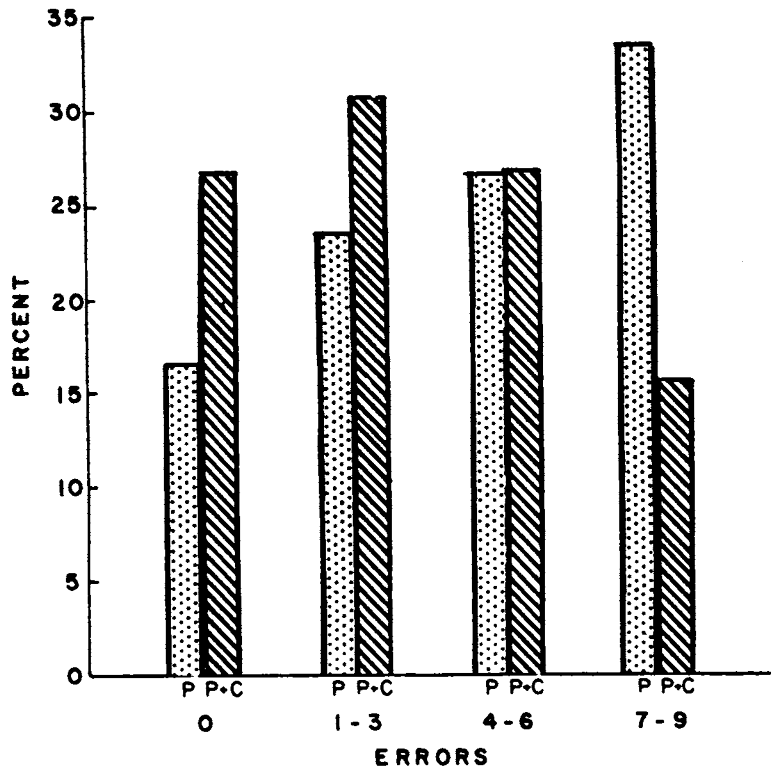


Figure 5. Frequency Distribution of Errors on Test 1.

significance of the obtained differences. If this is done, several of our more important comparisons lead to probabilities of less than .05 that the obtained differences occurred by chance.

Discussion

Although the results of the experiment are not as clear-cut as could be desired, they generally support our basic hypothesis that a training procedure that provides some opportunity for S to construct an answer in the absence of total cue support will be more effective than one in which 100 per cent stimulus support is always available.

The data appear to justify tentative acceptance of these specific conclusions: (1) on the first test (i.e., after half the total number of practice trials) the central tendency of the error scores for S's run in the P+C condition is lower than the central tendency of error scores for S's run in the P-alone condition (Mann-Whitney U Test); (2) of S's who made errors on the first test, a greater proportion of those run in the P+C condition than of those run in the P-alone condition made no errors on either Test 2 or Test 3; (3) for total error scores on all the tests, a greater proportion of cases of zero errors occurred in the P+C condition than in the P-alone condition.

Taken together, these results suggest the conclusion that a training sequence that incorporates some unprompted practice trials will be more effective than one that does not. The results of the Cook experiments that showed the superiority of the prompting procedure, of course, are not questioned, because of differences in the experimental procedures. Cook did not investigate a "pure" prompting condition, since unprompted trials were interspersed throughout his training sequence. On the other hand, the present experiment did not include a "pure" unprompted procedure, such as Cook's confirmation condition. The results of Cook's experiments and the present one taken together indicate that while an unmixed prompting sequence is superior to an unmixed confirmation sequence, it is less effective than a procedure in which both types of trials occur.

The question arises, of course, as to whether some arrangements of prompted and confirmed trials in the sequence might be more effective than others. It is the authors' hypothesis that a sequence in which the frequency of occurrence of confirmation-type trials increases progressively from very few early in learning to many late in learning (as implied by the notion of "vanishing") will be more effective than sequences in which (1) confirmation-type trials progressively decrease from many early in learning to few later, (2) the frequency of confirmation-type trials remains constant throughout learning (as in the confirmation condition of the present experiment), or (3) there is an abrupt shift from all-of-one-type trial to all-of-the-other-type trial. A systematic test of this hypothesis, involving varying "densities" of confirmation-type trials and varying rates of increase and decrease of confirmation-type trials, awaits future investigation.

The results of an experiment by Kaess and Zeaman (1960) provide evidence contributing to an interpretation of the present findings and those of Cook. Kaess and Zeaman, using multiple-choice items on a Pressey-type punch board, found that negative knowledge of results (i.e., information that a response is incorrect) tends to detain the reduction of errors on subsequent trials. Relating this to Cook's experimental design, it may be noted that a S in Cook's confirmation condition had less than one chance in 100 of making a correct response on the first presentation of a given stimulus, since he could only guess at what the code-line pattern should be. In other words, S very likely received negative knowledge of results. S's in Cook's prompting condition, on the other hand, received no negative information.

Thus, the design of the present experiment was such that a subject in the P+C condition had made three forced correct responses to each stimulus before being required to produce a response that was not forced. The chances of his responding correctly on the first "constructed" response trial, then, were considerably greater than the chances of a S in Cook's confirmation condition. Or, put another way, the probability of a S obtaining negative feedback was much lower in our experiment.

The obtained result of the present experiment was predicted on the basis, stated earlier, that a sequence of prompted and unprompted trials not only insures practice of the correct response, but in addition provides practice in the criterion situation of responding without prompting. If this analysis is correct, the implication for training and education is to have the learner respond to the minimum value of cue strength sufficient to insure a correct response (cf. Lumsdaine, 1959b). Future experimental work should be addressed to an examination of the validity and generality of this principle, particularly as it applies to the development of auto-instructional programs.

CHAPTER 25

SELF-PACING VERSUS AUTOMATIC PACING IN PAIRED-ASSOCIATE LEARNING[1]

Leslie J. Briggs

Introduction

Many variations of the paired-associate form of arranging learning conditions have been employed by experimentalists and laymen alike. The goals of the learning efforts, as well as the materials, methods, procedures, and apparatus, have varied so much, in fact, that the general term "paired-associate learning" has lost much of its former specificity in meaning. But the term still has some usefulness in describing the learning situation in which: (a) rather brief pairs of terms are to be learned and associated, and (b) the method of presentation and practice employs only the stimulus and response terms themselves, to the exclusion of context phrases, stimulus supports, and other words such as are found in reading prose or in continuous-discourse programs.

When the experimentalist employs the paired-associate method, he normally attempts to impose certain sequential and temporal controls over the subject's procedure during learning, and he paces the exposure of the total successive trials. Such control tends to "standardize" conditions for all subjects and all trials, and thus tends to assure that whatever the factors in the practice situation that promote learning, they are available to an equal extent for all items, all trials, and all subjects.

However desirable this form of control may be, it does not encourage variation in effort and learning method for items of varying difficulty, nor does it adjust the materials and procedure as learning progresses. Whether the experimenter provided prompting or confirmation to aid the subject, the total conditions have not necessarily been designed solely to maximize learning, but to test single variables or points of theory under standard conditions.

Nevertheless arrangement of materials in a form for paired-associate learning is often a practical way to achieve learning of certain "real-life" materials,

[1]The experimental work reported in this chapter was conducted as an in-service project at the former AFPTRC Maintenance Laboratory. A brief report on the study was previously given limited circulation as an unpublished Laboratory Note (Briggs, Plashinski, and Jones, 1955). (See also Footnote 1 to Chapter 23.)

and experiments in this form of learning are therefore appropriate both to discover optimum conditions for paired-associate learning and to evaluate some of the specific single factors in learning that may, in practice, be employed most effectively in a continuous-discourse program or some other arrangement of material.

The layman, whether teacher, pupil, or adult learner, often employs a paired-associate practice condition to learn things he needs to know. Since his purpose is not experimental or theoretical, he often adjusts his procedure to suit his requirements. Thus he does not rigidly follow either a prompting or a confirmation procedure for all items or all trials, nor does he insist upon responding to all items during each trial. He may "peek" at the response term for difficult items, and he may skip the confirmation term for items on which his confidence in his response is high. Further, he may sort his pairs as he goes along, eliminating from the series the pairs already learned.

From the viewpoint of practicality, the layman designs for himself a very flexible program characterized by: (a) employment of stimulus supports not in the paired terms themselves; (b) variation, by item or trial in prompting and confirmation, or neither; (c) varying frequency in accordance with difficulty; (d) dropping out items when mastered in order to concentrate effort on remaining terms; and (e) reviewing at what he considers appropriate intervals. It is the above program characteristics that are imitated in the more elaborate teaching machines; but seldom does the average machine or program achieve this total array of flexibility in procedure and material.

But we are here to praise experimentation, not to condemn it. Experimentation has as its goal the discovery of the effectiveness of the components of the complex learning process, and the determination of how interactions among the simple components of learning function to achieve success in mastery. Eventually, experimentation may progress to achieve synthesis of optimum total learning conditions, but in the meantime, programmers will proceed to put common sense and knowledge of simple variables together in the construction of relatively effective programs.

Purpose of the Paper

The purposes of this paper are briefly: to describe an apparatus that provides more flexible conditions for paired-associates learning than does the familiar memory drum; to enumerate some of the specifics of the conditions and variations thereof that it offers; and to present results concerning one aspect of its use, namely, the capability for either self pacing or automatic pacing.

In a companion paper (Chapter 23), the author has presented other data from experimental use of the same device. Together, the two papers report the results of much of the experimental use of the device, and then discuss the results in terms of learning issues of current interest, for the experimentalist and the programmer of automated instructional materials.

Development of the Subject-Matter Trainer

The Subject-Matter Trainer, as its name may imply, originally was developed to serve a practical purpose, namely, the learning of technical subject matter (in the original instance as required in certain Air Force training courses). As with many devices designed for practical use, its employment in the classroom was not deferred until the specifics of its usefulness were studied and verified in the laboratory. Rather, laboratory, classroom, and field experiments were conducted almost simultaneously. It was believed that the device could benefit training efforts in its original form, used in accordance with the dictates of judgment, while improvements in its use and changes in its design could later be effected as results from laboratory experiments accumulated. This belief was based on consideration of the general learning principles that guided the design; such as active participation, much guidance early and less guidance later, reinforcement, etc. The total project, intended to achieve a many-faceted evaluation and improvement over the specific learning conditions offered by the device, was terminated by the "discontinuation" of the Air Force Personnel and Training Research Center. Terminated at the same time were efforts to build self-instructional devices for more complex skills (Briggs and DuVall, 1957; Briggs, 1959), and a full-scale project on self-instructional techniques initiated by Gagné and Lumsdaine (Lumsdaine and Glaser, 1960, pp. 257, ff).

Nevertheless, a number of single-variable experiments were conducted to answer specific questions regarding effective use of the device, and several classroom tests and one field try-out were accomplished. Several of the laboratory studies yielded data that were never published, but were employed as guidance in classroom use. Some of the results of these unpublished studies will be mentioned in later portions of this paper.

The results on pacing, presented in this report, were first presented in mimeograph form by Briggs, Plashinski, and Jones (see Lumsdaine and Glaser, 1960, p. 591). A study by Irion and Briggs (1957) yielded data presented in Chapter 23. Results from a classroom experiment have been presented by Briggs and Besnard (1956). A field try-out of the device has been reported by Mayer and Westfield (1958), and two other uses were made in classrooms of the Air Training Command at Francis E. Warren Air Force Base and Chanute Air Force Base, with the assistance of J. Jepson Wulff and Ernst Z. Rothkopf.

The decision to design the Subject-Matter Trainer, appropriately enough, arose from personal observations of classroom procedures. Specifically, it was observed that instructors in maintenance courses for electronic technicians spent an excessive amount of class time in attempting to teach each student to recognize and name properly some 50 major components of an electronic system (Besnard, Briggs, Mursch, and Walker, 1955). Some students learned these quickly after the instructor had pointed to each object a few times while pronouncing its name. However, many students could not identify the components correctly, even after the instructor repeated the procedure daily for several weeks. The observers therefore undertook to develop a means whereby individual students could learn the identifications at their own rates by active practice, thus saving valuable class time spent in the daily "reviews" which interfered with other required instruction.

The resulting Subject-Matter Trainer is described in more detail in a number of sources (Besnard, Briggs, Mursch, and Walker, 1955; Besnard, Briggs, and Walker, 1955; Briggs, 1958). Essentially this device is an electromechanical apparatus that displays up to 20 response terms together and continuously on a response panel. Thus, for a series of 20 "paired-associates," the stimulus terms are presented much as in a memory drum, but the 20 response terms (which may be pictures of components) are presented together, making a 20-choice "multiple-choice problem" of a "recognition" nature. Thus the display is of a sort that does not fall clearly within either the traditional memory drum or the usual multiple-choice arrangements.

Perhaps the most unusual feature of the device is the provision of multiple "modes of operation," thus offering a more flexible range of practice conditions. A master control permits selection of the desired condition, which can be the same for all trials, or changed as learning progresses. These conditions vary as to use of prompting versus confirmation and in the number of attempts (errors) permitted on each item under the various confirmation conditions. Thus overt errors may be restricted to one per item in one mode of the confirmation condition, while multiple overt errors are possible in another condition.

Under the self-paced prompting condition, further defined below, the subject operates only two control buttons, one called the "stepping switch," to present the stimulus item in the window when desired, and the other the "quiz switch," to ask the machine to indicate the right answer. The machine responds to this request by activating a small light located next to the one response term of the 20 that is correct. In automatic-pacing operation, an electrical timer, rather than the "stepping switch," controls the interval of stimulus exposure.

Under the various self-paced confirmation conditions, the subject also uses the stepping switch to present the stimulus items when he desires them. He views the stimulus item, and he then presses a response button next to the item among the 20 on the response panel that he believes is correct. If his choice is correct, the green light next to the correct response term glows, providing confirmation. If his answer is wrong, under one condition a buzzer sounds and he must try again; under another condition, permitting only one overt error, the buzzer sounds but at once the green light adjacent to the correct answer glows, thus informing him he was wrong, but also indicating the right answer automatically.

A total of seven modes of operation is provided, (actually more, as all conditions permit either automatic or self-pacing). Only two will be described here, as only these two were employed in the study to be reported. Other modes are described further in Chapter 25 and in other references previously cited.

Two Modes of Operation Employed in the Experiment

Prompting Mode. In earlier papers this mode was called the "quiz" mode, because the subject presses the "quiz" button to request the machine to show him the correct answer. Since this is contrary to the normal connotation of a quiz as a test of the subject rather than a request by him for information, the term prompting mode is used here. However, "prompting" takes on a rather unusual meaning in this context. In the Subject-Matter Trainer the prompt, or right answer, is given to the subject as previously described, but the subject is not then required to pronounce, write, or otherwise produce overtly, the answer as in the procedures of prompting used by Cook (1956; 1958), by Angell and Lumsdaine (1960), and in most paired-associate or programming experiments. Rather, he only observes the answer indicated, and tries "covertly" to learn and remember it. Time permitted subjects to glance several times at both the stimulus term and the prompt term. It is this absence of a requirement of an overt response (to copy or repeat the prompt) that makes prompting of special significance here.

The prompting mode described here does not guarantee that the subject makes no covert anticipatory responses before he presses the quiz switch and views the answer (prompt) term indicated by the green light. We thus have the possible ambiguity that one subject may first view the stimulus term, then covertly try to construct an answer or to choose it (without overt indication) from one of the 20 on the response panel. In such an event, the presentation of the "prompt" by the machine may in effect become a confirmation (or a correction) for the covert choice made. However, subjects were instructed to press the quiz switch at once, and thus to refrain from trying covertly to select an answer before viewing the prompt term. Observation of subjects led to the belief that most subjects follow this instruction, especially when the prompting mode is used on the first learning trial.[2]

Single-Try Confirmation Mode. Again, the mode is re-named, thus differing from terms used in earlier writings. This confirmation mode first presents the stimulus term, after which the subject must try to recognize and select the right answer. Only one attempt is permitted. If the answer chosen is correct, the green (confirmation) light glows, and the subject may go on to the next item when he wishes. If the answer is wrong, this is indicated to the subject, but also at once the green light glows next to the answer that is correct. To make sure the subject recognizes the correction offered by the device, the subject is required then to press the response button next to the correct answer indicated before he can activate the stepping switch to have the next item presented. Thus in the case of a

[2]To clarify a matter of terminology, both the procedure described above and prompting as employed by Cook (1958) depart from one general connotation of "prompting." On the stage, a prompter gives a prompt (reads the next line) only when the actor tries to perform, but hesitates, or when he starts to perform an error. In the same context, a cue to enter the stage is given the actor before the actor tries to decide for himself whether to enter. The actor is trained to covertly anticipate the cue, but to wait for it before performing.

wrong response, several things happen. First, the buzzer signals an error. Second the right answer (correction) light glows. Third, the subject must press the right answer key, the green light remaining on. Fourth (at least in the self-paced condition), the subject is permitted to take time to re-study the stimulus item, and the indicated right answer, before going on to the next item. (In automatic pacing, the time taken to make the response determines time remaining to view the stimulus and the correction terms.)

The Pacing Experiment

Purpose

In early use of the Subject-Matter Trainer under self-paced conditions, no instructions were given subjects as to the benefits of either fast or slow pacing. They were simply told how many minutes of practice would be allowed, and they were shown how to operate the device. Since an observer was present making records of right and wrong responses, trial by trial, the subjects tended to proceed very slowly, apparently desiring to avoid making errors. The experimenters formed the opinion that under those conditions, the low frequency of responses and small number of trials were resulting in less effective use of practice time than would result from a more rapid pace. They concluded that for further experimentation, encouragement must be given to subjects to work more rapidly and not to regard avoidance of errors as especially important on practice trials.

In order to determine what instructions to give the subjects in regard to pacing when total practice time was controlled, it was decided that experimental data were required that would reveal the trade-off values of high frequency of trials versus attempts to avoid error, especially considering that in the absence of stimulus supports some error was unavoidable in early trials in all but the prompting mode of operation.

Brief experiments were therefore undertaken to study the effects (a) of various verbal instructions given subjects under a self-paced condition, and (b) of various time intervals for automatic pacing of the practice. These two lines of investigation, conducted by the writer, with assistance at various times by Guy G. Besnard, Daniel Plashinski, and Duane L. Jones, resulted in establishing the verbal instructions that appeared to produce the best learning results for self-pacing, and in deciding on an optimum electrical-timer setting for automatic pacing. The next step was then to determine whether either of these optimized procedures was superior to the other. It was considered that results from this determination should be used in deciding the condition for the later prompting versus confirmation study to follow (reported in Chapter 25).

Procedure

Two groups of Airman subjects, matched on the basis of scores from the Airman Classification Battery, were given practice on a 20-item paired-associate problem that required associating stimulus items consisting of names of men, (such as Art, Jim, Robert, Guy, etc.) with response items consisting of line drawings of various shapes representing real and fictitious types of electron tubes. All 20 response terms (shapes) remained visible on the response panel throughout practice and testing, while the stimulus terms appeared one at a time.

After the practice period and before the test trial, a new response panel was substituted in which the relative location of the 20 response terms was "scrambled." (On both the practice panel and the test panel, the 20 drawings were arranged in regular rows and columns, each drawing appearing adjacent to the corresponding green signal light and the response button. The location of the "scrambled" arrangement was determined by an electrical random-assignment circuit, which also served to assure that the device would always give accurate feedback regardless of how often response terms were "scrambled." Notches concealed along one edge of the metal response panel on which the 20 items were mounted served to show the experimenter where to mount each item in preparing a new scramble, and to keep feedback functioning accurately.)

The effects of scrambling had been the subject of prior experimentation that governed the decision to "scramble" only between the final practice trial and the test trial. Implications of this unpublished experimentation are discussed in a later section of this report. Based on this prior experimentation, subjects in this experiment, in both the self-paced and automatically-paced groups, were instructed that a scramble would take place before the test trial, and that they should thus avoid dependence on the position of the response terms in the 20-item array during learning trials.

Both groups of subjects were first given one trial in the prompting mode. Stimulus items were presented in a window at a constant rate of one each 13 seconds. This initial trial was automatically paced for both groups, to provide the same "start" in the learning. Upon the appearance of each stimulus item, the subject at once pressed the quiz button, whereupon the green light next to the correct answer was activated. The subject glanced at the two terms as rapidly as he wished while waiting for the next stimulus to appear. At a rate of 13 seconds per item, a trial lasted 260 seconds (± 1.5 sec.).

Following the initial prompting trial, as described above, all subjects were given 13 minutes of practice in the single-try confirmation mode, described earlier. One group was self-paced and the other was automatically paced at the 13-second rate. Only one overt response was permitted before the confirmation (or correction) was given. But after a wrong response, it may be recalled, the subject was required to acknowledge the correction by pressing the button next to the correction item indicated by the glowing light. (Only occasionally, in the automatic-paced condition, did a subject take so long for his response that the next stimulus was presented before the subject could make this second overt response after correction. The pacing of the initial trial served to give a frame of reference for judging elapsing time.) For both the self-paced and the automatically paced groups, an error response caused a buzzer to sound and a red light to flash, and at once activated the green correction light. (This double signal for error was included in this study to help subjects use time effectively; this point was made in the instructions, which were intended to avoid interpretation of these signals as punishment. For untimed experimental practice conditions, either or both of these error signals could be deleted by setting switches.)

For subjects in the self-paced group, of course, the number of confirmation trials completed varied. Subjects in the automatic-pacing group received three confirmation trials totaling 13 minutes.

At the close of the 13 minutes in the confirmation mode, all subjects were given the test trial, paced at 13 seconds, with nq knowledge of results. The number of items correct out of 20 was recorded for each practice trial and for the test trial.

Results

Of primary interest are the mean scores for the two groups on the test trial (Table 1). The difference between 10.45 for the self-paced group and 10.36 for the automatically paced group does not approach statistical significance ($t = .255$). Thus no difference between effectiveness of the two methods of pacing has been demonstrated by this comparison. It is conceivable, however, that were the

TABLE 1

Number of Correct Answers per Confirmation Trial

Self Pacing				Automatic Pacing			
Practice Trial	N	M	SD	Practice Trial	N	M	SD
1	38	8.82	4.79	1	39	7.15	3.57
2	37	9.30	3.89	2	39	9.30	4.36
3	30	11.10	4.77	3	39	11.33	4.82
4	26	11.77	4.96				
5	13	14.92	4.75				
6	5	15.00	5.93				
7	3	18.00	2.83				
Test Trial	38	10.45	5.35	Test Trial	39	10.36	5.71

automatically paced trials varied during practice to allow more time per item on early trials and less time on later trials, the differences might have been in favor of automatic pacing. This could be determined by another study. However, the present data suggest that with self-pacing, the use of pre-tested instructions to subjects that encourage them to try to use their own best rate, which they may vary throughout the practice trials, results in as much learning as in fixed automatic pacing at a rate considered to be optimal.

Of secondary interest is the indication that it takes a while for subjects to adjust to automatic pacing. Table 1 indicates that whereas the automatic group was poorer on confirmation trial 1 ($t = 5.4$), this difference disappeared subsequently. The t ratios for trials 2 and 3 are .00 and .613, respectively.

A third point of interest in Table 1 concerns the uniform increase in mean scores from one confirmation trial to the next, for both groups. The automatically paced group means progress regularly, with a drop of about one point on the test, which is not surprising considering scrambling and the absence of feedback on the test trial. But the mean scores for this group do not reveal an aspect of individual differences shown for the self-paced group in Table 1. Note that under self-pacing, based on pre-tested verbal instructions on how to benefit from self-pacing, there is yet great variance in the rate of pacing. One subject did not complete trial 2; eight subjects did not complete trial 3. But 26 subjects completed trial 4, and three subjects completed 7 trials. Note also that the means increase as the number of completed trials increase. Unfortunately, the original data have not been kept, and we cannot determine now the test trial scores associated with these varying rates of self-pacing. We might suspect that the fast pacers were fast learners, but the extremely fast pacers, who may have ignored the caution not to depend on position effects on the answer panel, may have succumbed to the effect of scrambling and made poorer test scores. On the other hand, those completing the most trials may have begun slowly, and learned well, thus speeding up while still performing correctly on later trials and the test trial. At any rate, the raw individual data are now permanently lost, and thus a lesson is pointed up on preservation of data for the future.

Discussion

For the conditions of learning employed in this study, no evidence is produced in favor of fixed automatic pacing, as compared to self-pacing preceded by verbal instructions found by trial to have improved pacing effectiveness over a no-instruction procedure employed earlier. This is not to say that automatic pacing that changes rate throughout practice would not prove superior to self pacing. [In this connection, cf. findings discussed by Maccoby and Sheffield in Chapter 5.]

These findings are for conditions including a single trial in a prompting procedure (paced), followed by a set period of practice in a confirmation condition. The materials employed familiar terms as stimuli, but required response learning of a perceptual-recognition nature, as well as association-forming between the paired terms. Any logical basis for the pairing would have to be developed implicitly by the subject, in the form of such a self-constructed stimulus support as "George goes with the short, fat, electron tube; George Smith is fat; remember this." Thus the materials employed resemble other paired-associate experimental tasks in that there is no context or stimulus support for the learning, other than such self-constructed ones, and the effects of training must depend largely on direct prompting and feedback, with varying frequency of response.

Whether the results on pacing are applicable to determination of pacing method for other forms of material, such as continuous-discourse programs, is uncertain. The results regarding pacing are for a condition of confirmation employed after one prompting trial. Although the study did not attempt to determine an optimal ratio of prompting to confirmation, it may be assumed that the scores obtained are higher than they would have been had the first trial been a confirmation rather than a prompting trial (cf. Cook, Chapter 21, 22).

It is desirable to recognize that the device employed in the study combines some features of presentation common to both the memory drum and a multiple-choice test format, and that both confirmation and prompting procedures contributed to the results. The form of response is a selection of one from 20 response choices, a rather unusual condition. However, the actual response-making process could well be "constructed," not "selected." The context in which a response is registered, and the mechanism for recording the response, do not rule out implicit mediating responses of a self-constructed nature.

To return to the matter of the scrambling of the answer choices, it may be remarked that early experience with the Subject-Matter Trainer appeared to yield some evidence both for and against "scrambling" on every practice trial as a means for increasing learning efficiency. That is, some subjects appeared to depend upon the position of the response term as a form of position-stimulus support, so that they improved with each trial but also succeeded in responding in late trials without this cue, and thus performed well if the panel was unexpectedly scrambled.

Other subjects who appeared to be using this same position factor as a support, failed almost completely when the panel was unexpectedly scrambled. To the extent that the "position habit" helped make right responses, which were confirmed or reinforced, the fixed position contributed to learning. But to the extent that the subject responded primarily in terms of position, the criterion learning was not taking place [cf. Wulff and Emeson, Chapter 30].

Although no clear experimental evidence then existed, the writer early adopted the practice of permitting the subject to practice with a fixed panel, with no mention of effect of position during early trials; then he warned the subject not to depend on position. After the warning and a trial or two, the subject was then required to continue practice with a scrambled panel. Although this procedure apparently worked best in practical learning situations, in subsequent experiments the verbal instruction was given at the outset to avoid dependence upon position, in order not to inject an additional change of condition during the experimental trials. The familiar concept of "position habit," often considered a negative factor in learning, thus appears potentially usable in this device as a stimulus support [cf. Chapter 30].

Summary

The experiment reported here involved two groups of subjects. Both groups were first given a paced trial in the prompting condition, followed by 13 minutes of practice in a confirmation condition. One group was allowed to determine its

own rate of pacing, each person setting his own rate at all times during the 13-minute period. Another group was paced automatically at a rate of 13 seconds per item.

A test was given under a pacing condition at the close of practice. No knowledge of results was supplied during the test, and location of answer terms in the 20-item array was scrambled, thus changing the locations so that success on the test could not be attributed to "position learning."

Results failed to show one group superior to the other; but this finding does not rule out the possibility that variable automatic pacing, slow at first, and faster as learning trials progress, could be superior to self-pacing. Results are discussed in terms of the conditions employed in the experiment, and in terms of earlier experimentation with the device.

CHAPTER 26

INTERPOLATED MOTIVATIONAL STATEMENTS WITHIN A PROGRAMMED SERIES OF INSTRUCTIONS AS A DISTRIBUTION-OF-PRACTICE FACTOR[1]

William J. McGuire

Introduction

A previous study (see Chapter 14) investigated the learning effect produced by introducing a symbolic reward sequence within an instructional film. The film taught techniques for performing at a rotary pursuit task by depicting proper and improper techniques. Symbolic rewards were introduced by showing the recording clocks registering good or poor scores. After predesignated instructions, the actor was shown to be getting a good score or a poor one, depending on whether he was adopting or violating the designated techniques. The Ss' scores (on extent of adopting the recommended techniques) showed that the symbolic reward sequences did have an effect; i.e., specific techniques were better learned to the extent that they were depicted close to a reward sequence. However, the reward had a proactive as well as retroactive benefit for various reasons (cf. Chapter 14). This "spread-of-effect" function (Thorndike, 1933) was interpreted as indicating that the vicarious reward sequences served to distribute practice, rather than serving primarily as a reinforcer.

The present study was designed to test further this interpretation of the role of motivational sequences within a programmed sequence of instructions. From the distribution-of-practice interpretation, several hypotheses were derived regarding the effects of interrupting the presentation of a series of instructions at predetermined points with interpolated motivational statements. It was hypothesized, first, that such interpolations would increase learning over the level attained in the uninterrupted presentation, but that an interpolated "silent" pause would be at least as efficacious as the interpolation of a pause of the same length filled with a motivational statement.

Secondly, it was hypothesized that the interrupted series would be superior to the uninterrupted series to the extent that the presentation conditions were such as would produce considerable reactive inhibition. (In the present experiment the condition varied was speed of presentation of the instructional material.) Indeed, it

[1]The experimental work reported in this chapter and the following one was, like the previous studies by McGuire reported in Part II, conducted at Yale University under Air Force contract AF 18(600)-35, monitored by A. A. Lumsdaine. (See also Footnote 1, Chapter 12.) The writer, who is now at Columbia University, expresses his indebtedness to J. J. Wulff, F. D. Sheffield, and G. A. Kimble for advice and suggestions on various phases of the work reported here.

was predicted that, as the conditions became productive of large amounts of reactive inhibition, the silent interruptions would become more efficacious for learning than the pauses with motivational statements.

Thirdly, it was hypothesized that the interpolation of pauses with motivational statements—and especially of silent pauses—would reduce serial-position effects as compared with such effects in the uninterrupted-presentation condition. This lessening of the serial-position effect was predicted as being particularly pronounced to the extent that the conditions of the instructional program made for large amounts of reactive inhibition (in this case, by the use of fast rate of presentation of the instructional material).

Method

Subjects and Materials

S's were 48 males, between the ages of 17 and 26 years. All S's had the same task, consisting of learning the names of nine mechanical parts taken from an automobile fuel pump and other equipment. The parts were selected to be dissimilar in appearance, except that all were flat and less than one and one-half inches in their longest dimension (to facilitate photography). The names were all four or five letters in length, and were not descriptive of the parts. The parts were photographed, both with their names and unlabeled. A set of nine 35-mm. slides, each showing one part and its name, constituted a teaching series; a set of nine slides, each showing an unlabeled part, constituted a practice (or test) series.

Apparatus

The slides were presented to the S's by a slide projector with an automatic change device. During the six teaching trials, each slide (consisting of a picture of the part and its name) was presented for either four seconds (in the Slow Rate of Instruction condition) or two seconds (in the Fast condition). On the six test trials —one following each teaching trial—each slide (consisting of an unlabeled picture of a part) was presented for four seconds. Between the end of one trial and the beginning of the next there was a blank interval of six seconds. In each trial the nine slides were presented in a different random order.

Design and Experimental Units

The rate of presentation was manipulated by showing each slide in the teaching series for four seconds to S's in the slow-rate-of-presentation condition, and for two seconds to S's in the fast condition. This differential rate was operative only in the learning trials, not on the final test trials.

The second independent variable, related to motivational induction, was varied in three ways. For one-third of the S's, the nine slides in each trial were presented without interruption (no pause). For a second third of the S's, a four-second interruption in presentation of the slides came after the third and sixth slide on each trial.

The interval was filled with a motivational statement by E, who commented alternately during these intervals in either of two ways: "Try to remember the name of each part shown in the slides so that you can get a good score in the final test," or, "If you try hard to memorize the name of each part you will probably get a good score on the final test." The remaining one-third of the S's were also shown the slides with these two interruptions in each trial, but the interruptions served more exclusively as "rest" pauses, with the E remaining silent.

Results and Discussion

The results of the present experiment are disappointing in the significance levels of differences. However, the predicted directions of the differences are sufficiently frequent to invite replication with a larger number of subjects.

The first hypothesis receives some support, as can be computed from the data in Table 1. Scores obtained in the combined pause-and-motivation and simple-pause conditions were higher than in the no-pause condition. The mean number of correct associations per trial, over the six trials, was 5.24 for the pause conditions and 4.32 for the no-pause series. (This difference, attains an .05 level of confidence for one tail.) On the other hand, the pause condition was, as hypothesized, just as good (in fact, trivially better) as the pause-plus-motivation condition, the means being 5.24 and 4.94 respectively.

TABLE 1

Mean Number of Names of the Nine Parts Correctly Recalled
Per Trial in the Six Conditions

	Type of Pause Condition		
Speed of Presentation	Uninterrupted (No Pause)	Pause Plus Motivation	Pause Only
Slow	5.99	6.45	6.34
Fast	2.66	3.43	4.14

The second hypothesis, dealing with interactions between speed of presentation and type of motivational pause, received support in the direction of the results, but the obtained differences do not attain conventional levels of significance. The two pause conditions were not significantly more superior to the uninterrupted condition at the fast presentation rate than at the slow as far as simple differences between means were concerned. Comparing the two pause conditions with one another, the pause-and-motivation group is slightly superior at the slow speed, and the pause-only group is slightly superior at the fast speed, but the predicted interaction is far short of the conventional significance levels.

The third hypothesis (derived like the first two from the postulate that interjected motivating comments derive their efficacy from distribution of practice) had to do with the serial-position effect. It was predicted that the relatively "slow-learning" items toward the middle of the list would be less pronounced in the pause plus-motivation-condition—and, especially, in the pause-only condition. Since this reduction in the serial-position effect would manifest itself to the extent that practice was massed, it was predicted that this differential in serial-position effect would be especially pronounced under the fast-rate-of-presentation condition.

The results depicted in Figure 1 bear out this prediction. It can be seen that with the slow rate of presentation there was little evidence of any serial-position effect, but at the fast rate of presentation the predicted differentials occurred. For the fast condition, the middle third of the list contributed 37 per cent of the errors

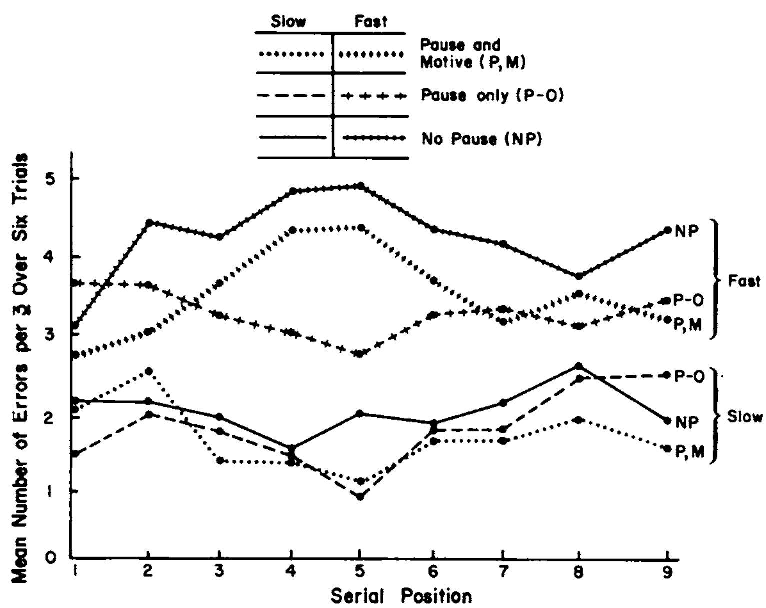


Figure 1. Comparative serial-position effects in the three motivational-pause conditions at slow and fast rate of presentation

in the No-Pause (NP) series, 38 per cent in the pause-plus-motivation condition (PM), and only 30 per cent in the pause-only (PO) condition. The variance among the means is significant beyond the .05 level. The only difference between pairs of means that attains this level of confidence is that between the no-pause and the pause-only conditions, which were predicted as being the most different in this regard.

Summary

It was postulated on the basis of previous findings that part, at least, of the benefits of interjecting motivational statements in a training film was related to the greater spacing of trials necessitated by such statements, and therefore that the interjection of a silent pause of the same length would be at least as efficacious in promoting learning and reducing the serial-position effect.

A modified paired-associates task was presented by means of a series of slides without interruption; a second group saw the series with two interruptions on each trial during which the E made motivational statements; a third group saw the series interrupted by silent pauses of the same length at the same points in the series. Rate of presentation was varied orthogonically to the pause condition. The S's were 48 males.

It was found that the two pause conditions combined were superior to the no-pause condition (at a borderline level of confidence), and that the pause condition was at least as effective as the pause-plus-motivation condition. Likewise, the pause conditions tended to show less serial-position effect than the no-pause condition, the pause-only condition showing significantly less serial-position effect than the pause-plus-motivation condition.

CHAPTER 27

AUDIENCE PARTICIPATION AND AUDIO-VISUAL INSTRUCTION: OVERT-COVERT RESPONDING AND RATE OF PRESENTATION[1]

William J. McGuire

Introduction

Mass instruction, even in face-to-face lecture situations, tends to reduce the learner's opportunities for participation in the learning process. The additional requirements for silence and reduced illumination when audio-visual aids are introduced into the situation further reduce these opportunities. It is generally agreed on the basis of theoretical considerations and the results of the reading-versus-recitation studies (Gates, 1917), that participation is conducive to learning. Hilgard (1953), for example, lists this principle as an uncontroversial generalization from human learning experiments. Hence, users of audio-visual presentation methods may be somewhat regretful about this side-effect of their procedure.

The present paper presents a theoretical analysis of the effects on learning of providing an audience with participation opportunities during an instructional presentation. This analysis points out that such opportunities can have deleterious as well as beneficial effects on learning and attempts to specify situational characteristics that will make one or the other class of effects predominate. The effects of participation opportunities are further analyzed with respect to whether the participation is covert or overt. In the former condition the student (S) is instructed to participate "mentally" or "to himself," and is given time to do so, but with his responses not perceptible to the instructor or to other S's. (Such responses presumably take the form of sub-audible verbal surrogates or incipient movements corresponding to the performance for which he is being trained.) With overt response conditions, on the other hand, the participation response made by S (or lack thereof) is perceptible, and usually consists of the precise performance for which he is being trained. The results of a small-scale study to test some of the hypotheses derived from this analysis are also presented.

[1]The experimental work reported in this chapter was previously reported in a laboratory note (McGuire, 1955b) on research conducted at Yale University, as were the previous studies by McGuire in Part II, under Air Force contract AF 18(600)-35, monitored by A. A. Lumsdaine. (See also Footnote 1, Chapter 12 and Footnote 1, Chapter 26.) The writer, who is now at Columbia University, expresses his indebtedness to J. J. Wulff, F. D. Sheffield, and G. A. Kimble for advice and suggestions on various phases of the works reported here.

Possible Beneficial Effects of Participation

Distribution of Practice

Introduction of participation opportunities, whether covert or overt, during the instructional session tends to distribute practice, since the S alternates between trying to master the instructions and trying to put them into practice. This advantage of participation (which was pointed out by Ruben-Rabson, 1941) is important when the instructional situation involves conditions that make for a deleterious massed-practice effect: difficult task, long session, homogeneous material, fast presentation rate, etc. (McGeogh and Irion, 1952, Ch. 5.) (On the other hand, it might actually be better to employ massed practice during early stages of learning in situations where the S's begin the session with a predominant tendency to make a wrong response, since this wrong response may be more quickly extinguished by massed training.)

Feedback

Both overt- and covert-participation conditions afford the S an opportunity to test the efficacy of his technique for mastering the task, while there is still time to change his technique if it proves inadequate. Such an opportunity is important to the extent that there is a likelihood that the S initially adopts an inefficient technique, the inadequacy of which he would probably not become aware without testing.

This feedback from participation also facilitates the S's identifying the part of the task in which he is doing most poorly and where, therefore, he ought to concentrate his effort. This opportunity becomes more important to the extent that the task is composed of sharply defined sub-tasks, differing in difficulty.

Motivating Disinterested S's

Participation conditions, at least overt, may increase the S's motivation to cooperate in the learning process, as has been suggested by Hovland, Lumsdaine, and Sheffield (1949). Such will be the case to the extent that fear of being observed not participating will motivate the S to make a greater effort to learn, or at least appear to be learning, the task. We might expect such a situation to arise, if the Ss have little intrinsic motivation to learn and think that there will be no other test of their learning except the extent of their participation (cf. Chapter 19); and particularly if they also fear the sanctions they expect if observed not participating. Such circumstances may be present in situations involving S's who are hostile to the instructor or to the materials being taught, as in certain cases of attitudinal indoctrination such as those termed "brainwashing." It seems likely that overt-participation procedures facilitate the learning of new attitudes in such situations, and some experimental corroboration is furnished by Janis and King (1954). But in situations where S's have a strong desire to learn for some intrinsic reason, they expect to be tested at the end of the practice session in any case, and have little fear of possible sanctions if observed failing to participate, then the added motivation deriving from this source should add little to their already existing motivation to learn. For example, Hovland, Lumsdaine, and Sheffield (1949) found that there was less difference between participation and non-participation procedures when all groups were informed in advance that a test would be given at the end of the instructional period. Similarly, Michael and Maccoby (1953) found evidence for doubting that any advantage of a participation procedure derives from its motivational effects under conditions of presumably high extrinsic motivation (cf. Chapter 18).

Response Generalization

Participation tends to allow practice with responses similar to those the S must ultimately learn to make. When the task being taught involves acquisition of a manipulatory skill, overt participation makes such direct practice possible, while in covert or no-participation conditions the S is largely restricted to practicing with only verbal surrogates of the manipulative responses. It is an accepted principle that the response during the learning session should be as similar as possible to those which must be performed in the eventual test situation (perhaps within limits imposed by the formation of reactive inhibition during practice). Hence, overt participation would seem to offer a decided advantage over covert- and no-participation procedures in the case of hard-to-verbalize manipulatory tasks. Yet empirical results are perhaps more striking in the extent to which they demonstrate that even covert "symbolic" participation is efficacious in producing mastery of motor skill tasks. For example, Vandell, Davis, and Clugston (1943) found that "mental" practice effected as much improvement in dart throwing and basketball shooting accuracy as did actual physical practice. Twining (1949) found that physical practice effected a greater improvement in skill at ring tossing than did mental practice, but that the latter also resulted in considerable improvement (cf. also Harby, 1952). Similar findings in connection with "mental" versus keyboard piano practice are reported by Rubin-Rabson (1941 a,b). Perry (1939) found "imaginary" practice superior to actual practice on a peg-board task. As opposed to this question of the efficacy of the covert procedure with motor-skill tasks, less superiority of the overt over the covert condition would be expected when the task being taught is itself a verbal one, since the covert-participation responses would then be sub-audible verbalizations—quite similar to the overt verbalizations to be learned.

Of these four possible beneficial effects of participation on learning, the first two would be equally operative in either overt- or covert-participation procedures, while the last two would be more operative in overt than in covert. That is, overt participation tends to offer advantages over covert when the S is poorly motivated to learn the task, or when the task involves skilled movements for which there are no closely corresponding verbal or incipient movement surrogates that will allow them to be adequately practiced covertly.

Possible Deleterious Effects of Participation

Motivational Effects

If the S is called upon to participate before he is able to do so to any considerable extent, he may be demoralized and left feeling inadequate and anxious.

As Taylor and Spence (1952) and Child (1954) have pointed out, a small amount of anxiety may have a positive effect on learning, but a high level of anxiety may elicit responses (such as hostility and psychological disengagement from the situation) that interfere with the desired learning. Such deleterious effects are likely to be produced if the participation session is introduced too early in the instructional session with a difficult task, and without forewarning the trainee that he will not do well in the absolute sense during the early participation trials. For example, it was found by Louise Krueger (1930) and by W. C. F. Krueger (1930b) that when a single participation interval, without prompting, is employed, it should be introduced at a late rather than early stage of the instructional session. This effect increases as the difficulty of the task increases or involves responses that are easily interfered with, and as the S's general anxiety level increases. Kimble and Wulff (cf. Chapter 16) have demonstrated the deleterious effects of excessive participation in a complex task.

Maladaptive Learning

There is also the danger that participation may strengthen incorrect habits, a danger that is present when incorrect responses are likely and may be quickly followed by correct, reinforced responses, resulting in the (delayed) reinforcement of the prior incorrect response. Since the efficacy of reinforcement declines rapidly as delay increases, the seriousness of this possibility diminishes rapidly as the participation pace becomes slower. Also, this danger is greater with overt- than with covert-participation procedures, since it is in the overt condition that the pressure to participate and consequently the likelihood of making wrong responses is greatest.

Interference with Reception

It is possible that if pressure to participate is great the S's preoccupation with participating might interfere with his reception of the instructional material. This danger is present to the extent that teaching and participation periods are extensively intermingled, the period allowed for participation is brief, and the duration of presentation of instructional information following the participation is also brief. Some evidence for this postulated disadvantage has been found. Jaspen (1950b) showed that overt participation may have a detrimental effect with quickly paced instructions, and Roshal (See Chapter 11) interprets his results as indicating the importance of allowing adequate time for participation. It is also possible that in the case of group learning situations certain kinds of overt (but not covert) participation will have the disadvantage that one S's responding may interfere with the learning of the others. This possibility was not involved in the present study, since the S's participated individually.

All three of these potential deleterious effects of participation would tend to be more operative with overt than covert participation. Furthermore, they are more pronounced under conditions of rapid information presentation.

Experimental Demonstration

An experiment was designed on the basis of the above analysis to demonstrate that under some conditions audience participation is conducive to learning, but under other specifiable conditions it interferes with learning, and also to show that an overt-participation requirement is more effective under some conditions, and covert under others. Since the beneficial effect of participation has often been demonstrated before, we deliberately designed the present situation so that the hypothesized detrimental effects would be accentuated.

Three of the four hypothesized beneficial effects of participation were minimized in the present study. Distribution of practice was rendered unimportant by employing a brief task with inter-trial rest; motivational effect was minimized by telling all S's in advance that they would be tested at the end of instruction; and problems in response generalization were reduced by using a verbal task that could easily be practiced covertly. A fourth possible beneficial effect remained in the situation, in that the participation provided better feedback to the S regarding the more difficult parts of the task on which he might concentrate his effort. Besides minimizing the benefits of any kind of participation, these experimental conditions largely eliminated the potential benefits of overt over covert participation (by minimizing the motivational and response generalization factors that could favor the overt condition, while leaving operative only the feedback opportunity that is provided equally by overt and covert).

On the other hand, the conditions of the experiment were set up to make the hypothesized detrimental effects of participation more pronounced. A homogeneous task was selected with the thought that this might maximize any interfering effect of increased test anxiety; likelihood of maladaptive learning was increased by requiring participation even in the early stages of learning; and interference with reception by a rapidly paced presentation. These conditions, further, were all such as would be more disadvantageous in overt- than in covert-participation conditions, and at fast rather than slow speed. The conditions of the present study were selected in terms of theoretical interest, and were not intended to be typical of actual mass, audio-visual instructional situations.

The two variables manipulated in the present experiment were: type of participation required (overt, covert, and none), and speed of instruction. An obvious hypothesis was that performance would be poorer (that is, the task more difficult) with the faster speed of instruction. The hypotheses of theoretical interest deal with interactions between the speed variable and the particpation variable. It was predicted that with greater speed of presentation the advantage of overt participation over no participation would decline; and furthermore that the advantage of overt over covert participation would be lessened with increased speed.

Method

Materials and Apparatus. All S's had the same task (the same as that used in the study reported in Chapter 26), consisting of learning the names of nine mechanical parts. A set of nine 35-mm. slides, each showing one part and its name, constituted a teaching series; a set of nine slides, each showing an unlabeled part, constituted a participation (or testing) series. The slides were projected on a transparent daylight screen, thus allowing the general room illumination to be bright enough to facilitate the S's doing the paper-and-pencil-participation task while viewing the slides. A slide projector with an automatic change device was used. The time intervals employed were as described in Chapter 26.

Design and Experimental Units. A 2 x 3 factorial design was used, with rate of presentation as one independent variable (varied in two ways), and type of participation as the other (varied three ways). Each S served in only one of the six conditions; since 48 S's took part in the study, eight served in each condition. The S's were assigned to conditions randomly and were run individually. All of the S's were males between the ages of 17 and 26. Their mean Stanford-Binet I. Q. score was 93; the S.D. of the distribution was 11.

Independent Variables. The rate of presentation was manipulated (in teaching trials only, not on test trials) by showing each slide in the teaching series for four seconds to S's in the slow rate of presentation, and for two seconds to S's in the fast condition.

Three categories were used for the second independent variable, type of participation—none, covert, and overt. In the no-participation condition, S's were told that they would be shown a series of nine slides, each picturing a part of the automobile fuel pump and its name; that they would be shown this series six times, the slides coming in a different order each time; that after all of the six teaching series had been shown, they would be given a test series consisting of pictures of the parts without their names, their task being to write down the name of each part as it was shown in this final series.

S's in the covert-participation condition were given similar instructions, and told in addition that after each teaching series there would be a practice series consisting of pictures of the parts without their names, as would the final test. The S's were instructed to try to think to themselves the name of each part as it was shown during these practice series.

S's in the overt-participation condition were given the same instructions as those in the covert, except that the S's in this overt condition were asked to try to write down on the test sheet the names of the parts instead of just trying to think of their names silently during each of the participation series of the practice session, just as they would have to do during the final test series.

After these instructions, questions raised by the S's were answered, and then they were shown the series under the conditions that had been described to them.

Results

Transformations. The dependent variable—amount of learning—was defined as the per cent of parts correctly named on the final, test trial. Several of the S's in the slow-rate-of-training conditions showed perfect learning after six trials, and two S's in the fast-rate conditions showed zero learning; the distributions were skewed and there was considerable heterogeneity of variance ($p = .08$ level using Bartlett's test). To correct for these factors, an inverse-sine transformation of the percentage-correct scores was made. (See Supplementary Table 1 for original and transformed data.) This transformation reduced the heterogeneity of variance considerably. (The heterogeneity of variance for the transformed scores was of a magnitude that might be expected to occur by chance 82 per cent of the time.) The results and significance levels in the rest of this paper are given in terms of these transformed scores. However, similar analyses of the original scores yield substantially the same interpretations (cf. also Norton 1952). Table 1 shows the mean performance in each of the six groups in terms of this transformation.

TABLE 1

Learning Scores in the Six Conditions
The score in each cell is the mean (based on eight S's) of the inverse sine of the proportion of correctly named parts on the final, test trial.

Presentation Rate	Participation condition: none	covert	overt
Slow	61	76	78
Fast	50	64	37

Overall Effect. As can be seen from Table 1, performance was indeed better ($p < .001$) at the slow-presentation rate; however, the overall contribution to the variance of the three participation treatments was not significant ($F = 2.32$). Table 2 presents separate analyses, A, B, and C, in which presentation treatments are compared in pairs. Table 2-A shows only a slight, non-significant difference favoring the overt over the no-participation condition; Table 2-B shows a significant difference favoring covert over no participation ($p = .05$), while Table 2-C fails to show a clearly significant difference ($p = .10$, two tails) between covert and overt participation conditions. These general differences between the participation conditions are, however, of little theoretical interest since, according to the hypotheses, one or another of the conditions might be superior depending on the specific speed parameters, which were chosen rather arbitrarily. Of theoretical interest are the interactions between participation and speed. The nature of these interactions will become clearer as we make more specific comparisons.

TABLE 2

Analysis of Variance for Transformed Scores Under Slow and Fast Rates of Presentation, Comparing Pairs of Participation Conditions

A. Comparison of overt and no participation

Source	s.s.	df	m.s.	F	p
Participation (overt versus no participation)	38	1	38	-	-
Rate (fast versus slow)	5278	1	5278	12.60	.001
(participation)(rate)	1910	1	1910	4.56	.05
residual	11723	28	419		

B. Comparison of covert and no participation

Source	s.s.	df	m.s.	F	p
Participation (covert versus no participation)	1667	1	1667	4.15	.05
Rate (fast versus slow)	957	1	957	2.38	.14
(participation)(rate)	4	1	4	-	-
residual	11244	28	402		

C. Comparison of overt and covert participation

Source	s.s.	df	m.s.	F	p
Participation (overt versus covert participation)	1201	1	1201	3.00	.10
Rate (fast versus slow)	5565	1	5565	13.88	.001
(participation)(rate)	1740	1	1740	4.34	.05
residual	11236	28	401		

Participation Versus No-Participation Effects. It was predicted that there would be less advantage of the overt-participation over the no-participation conditions with greater speed of presentation. As is seen in Table 1, this prediction was confirmed by the results. The mean learning score at the slow rate of presentation was 77 in the combined overt and covert conditions and only 61 in the no-participation group. At the fast rate of presentation, on the other hand, the learning score in the no-participation condition (50) and the average of the two participation conditions were about equal. The latter, however, is an average of two grossly dissimilar values (64 and 37); and the results are clearer if we compare the no-participation means with the covert and overt participation means separately. The experimental situation was designed so that potentially detrimental effects of participation would be negligible when the participation was covert. As predicted on this basis, the covert groups remained superior to the no-participation groups, the difference between means being 15 at the slow rate and 14 at the fast rate, both in favor of the covert condition (see also Table 2-B). The detrimental effect was predicted to increase with presentation speed in the overt-participation condition. This prediction is confirmed in that the overt participation group is considerably superior to the no-participation group at the slow-rate task (78 versus 61), but inferior (37 versus 50) at the fast rate (interaction of $p = .05$: See Table 2-A).

Overt Versus Covert Participation. The experimental situation was devised so that the potential advantages of participation would be equally operative in the overt and covert conditions, but the disadvantages more operative in the overt conditions. Furthermore, it was assumed that the disadvantages would be more important as the rate of presentation increased. The results are in accord with the consequent predictions. At the slow rate, the overt group performed only negligibly better than the covert (78 versus 76), while at the fast speed the covert-participation group was considerably superior to the overt (64 versus 37), the interaction effect being significant at the .05 level of confidence (see Table 2-C).

These results provide a confirmation of some of the hypothesized detrimental effects of overt participation and of the conditions most likely to make these effects sizable. In addition, several practical disadvantages of providing participation opportunities remain to be mentioned. In the first place, it should be noticed that when participation did produce superior learning (namely, at slow speeds), this improvement was achieved at the expense of doubling the length of the practice session, since the participation trials were superadded to the presentation trials. We might, of course, have required participation during the presentation trials, but this device would have maximized one of the detrimental effects of participation (namely, interference with reception of the instructional material) and lessened one of its beneficial effects (namely, feedback regarding places of poorest mastery). A second practical difficulty involved in providing opportunity for participation is the complications thus introduced into the training situation, such as the special lighting arrangements and the participation material that had to be employed in the present situation. Indeed, when it becomes necessary to provide participation materials for all S's receiving the audio-visual instruction, we might wonder if one of the major practical reasons for using audio-visual devices in the first place has not been negated.

Summary

A theoretical analysis of the advantages and disadvantages of participation over no participation during learning, and of overt over covert participation, was presented, together with a discussion of situational factors that would interact with the participation conditions. On the basis of this analysis, an experimental situation was devised in which it was predicted that the advantages of overt participation would decline as speed of instruction increased.

Forty-eight S's were shown instructional series of slides presenting pictures of mechanical parts and their names. Various groups had opportunities after each instruction series to participate overtly, covertly, or not at all in the learning task. Rate of presentation during the instructional series was varied between two and four seconds per slide.

The findings were in keeping with the predictions. Overt participation resulted in more learning than did no-participation at the slow rate of presentation, and less at the fast rate (interaction effect of $p = .05$), while covert resulted in more than did no participation at both slow and fast rates. At the slow rate, overt and covert participation were about equally effective, but at the fast rate, covert was superior (interaction effect of $p = .05$).

SUPPLEMENTARY TABLE 1

Individual Raw Scores of Number of Correctly Named Parts, and Transformation of These Scores (in parentheses) Using an Inverse-Sine Transformation, on the Final Test Trials

Overt participation		Covert participation		No participation	
Slow	Fast	Slow	Fast	Slow	Fast
9 (90)	7 (58)	8 (64)	5 (50)	7 (58)	9 (90)
9 (90)	3 (35)	9 (90)	9 (90)	3 (35)	6 (55)
9 (90)	5 (50)	8 (64)	8 (64)	3 (35)	6 (55)
7 (58)	0 (00)	9 (90)	9 (90)	9 (90)	5 (50)
9 (90)	0 (00)	9 (90)	4 (40)	8 (64)	2 (32)
9 (90)	5 (50)	9 (90)	3 (35)	8 (64)	2 (32)
6 (55)	6 (55)	6 (55)	9 (90)	9 (90)	7 (58)
8 (64)	5 (50)	8 (64)	6 (55)	5 (50)	2 (32)

CHAPTER 28

OVERT AND COVERT REHEARSAL OF FIFTY PER CENT VERSUS ONE HUNDRED PER CENT OF THE MATERIAL IN FILM LEARNING[1]

Joseph H. Kanner
Richard L. Sulzer

Introduction

The present study was designed to supplement the information provided by previous studies with respect to two major questions. First, what is the comparative effectiveness of different forms of active recitation in film learning? And second, what is the effect of covering only a portion of the instructional points in recitation sessions, as compared to covering all of the points?

Background

In a first experimental study of recitation periods interspersed between sections of a film, Hovland, Lumsdaine, and Sheffield (1949) demonstrated that oral recitation by students viewing a film-strip on the phonetic alphabet produced recall scores superior to those obtained by devoting equivalent training time to a conventional, passive review (see Chapter 1). The implementation and results of this study and the later study by Michael and Maccoby [see Chapter 18] have particular relevance to the present investigation.

Some of Hovland, Lumsdaine, and Sheffield's (1949) chief results were as follows: (1) The "active"-review (oral participation) group achieved superior test scores; (2) The superiority of the "active"-review group was greatest for the

[1]This study was initiated as an in-service project of the Audio-Visual Research Division at the U. S. Air Force's Human Resources Research Laboratories. The study was jointly designed by A. A. Lumsdaine and J. H. Kanner. The latter, as principal investigator, was responsible for the implementation and conduct of the study and for initial analysis and reporting of its results. The field tests were conducted in 1952-53 at Sampson Air Force Base by Francis H. Thomas and Andrew H. McClure. Further analysis and preparation of a final report, subsequent to Dr. Kanner's leaving the Air Force program in 1953, was accomplished by Dr. Sulzer under the direction of Dr. S. M. Roshal. The present chapter, aside from relatively minor editorial changes, reproduces the final report prepared by Dr. Sulzer, which was approved by Dr. Kanner for publication. A substantially identical version of the paper was reproduced for limited distribution in mimeographed form (Kanner and Sulzer, 1956).

more difficult items; (3) "Test-announcement" sub-groups attained higher scores than "no-announcement" sub-groups, but the difference due to presence or absence of the test announcement was much greater for the "passive"-review condition. (Seemingly, the students in the oral-practice condition did not require such additional motivation to learn as may result from a test threat.) The authors concluded that substituting recitation for conventional review may increase learning from films, particularly when individuals are poorly motivated. One possible theoretical analysis of the learning process that would account for this effect would hold that, for learning to occur, explicit practice in associating the stimulus and response processes related to the letters and the code words is necessary. Instructions to call out the answers, or announcement of a test, would serve to promote practice, and hence to increase learning. The effectiveness of participation procedures may not, however, be limited to the specific material that the students are instructed to rehearse overtly. Periodic recitations may tend to have a generalized or incentive effect on the learner's implicit practice of additional instructional material, although practice of these additional items is not formally required.

The second study that gave impetus to the present investigation was performed by Michael and Maccoby, as reported in Chapter 18. The results of their post-film tests showed that, when "feedback" of the correct answers was provided, either overt or covert practice produced considerable improvements in learning; and that without provision of the correct answers, after the subjects had formulated their own answers in the review-practice sessions, participation tended to result in superior test performance as compared to the straight film, no-review condition. This increase in learning was ascribed to the effects of directed practice and not to changes in motivation, since the increases obtained in learning were confined to the specific instructional items covered in the practice sessions. The other major new finding coming from Michael and Maccoby's study was the apparent equivalence of covert and overt practice for factual items. Otherwise, the results parallel those of the Hovland, Lumsdaine, and Sheffield experiment in showing that participation improves test scores, although different subject populations were sampled, different specific "active" review procedures were employed and different learning tasks were used in the two experiments.

In a later experiment conducted by Levine [see second section of Chapter 19], it was found that some carry-over of learning gains from "active" review to additional items took place when "low-motivation" classes were tested. With groups previously warned that a test would follow the film, however, no evidence for a "motivational effect" of rehearsal was obtained. The film on world maps appeared to be intrinsically less interesting than the Michael and Maccoby film on atomic warfare, and the combination of an uninteresting film and absence of a test announcement was designed to produce a low-motivation condition to make it possible to detect any motivational carry-over effects that might occur under such conditions.

Purpose of This Experiment

The present study was designed to gain additional information concerning the effectiveness of active-review procedures, along two dimensions: first, partial versus complete coverage of the items in rehearsal sessions; and second, comparison of more alternative kinds of rehearsal procedures than have been compared previously. It also provided a replication to check the possibility that Michael and Maccoby's findings (on non-carry-over to non-practiced material, and on the effectiveness of "covert" as opposed to "overt" practice) might hold only for certain classes of learning tasks, or only for certain practice procedures. Furthermore, only overt oral practice had been studied previously with a paired-associates task (learning the phonetic alphabet), and only overt written practice with substance learning of the kind used by Michael and Maccoby. The present study used both written and overt oral practice as well as covert practice with paired associates.

Experimental Design and Conditions

As in the study by Hovland, Lumsdaine, and Sheffield (1949), the standard film presented letters and their phonetic equivalents singly in alphabetical order in three blocks of six letters each, A-F, G-L, and M-R, and a final block of eight letters, S-Z. Rehearsal or review sequences were inserted after each block of letters and consisted of a single repetition (taking the form of one or another of the treatments outlined below) of each letter in the preceding block.

Five instructional treatments, each including the standard phonetic alphabet film, were compared in this experiment under conditions of complete coverage of the list of instructional items (cf. Table 1). Overt written practice (B), covert practice (C), and covert practice with "feedback" of the correct answers (D), similar to Michael and Maccoby's conditions, were compared with the overt oral technique (A) used by Hovland, Lumsdaine, and Sheffield. (The latter, used in a class group, provides a mixture of prompting and feedback to students from hearing the responses of other members in the class.) "Passive" review (E) and straight-film, no-review (F) treatments were also included.

To assess the effects of partial coverage of the instructional items in review sequences for the four active-review treatments and the passive review treatment, additional classes of students were given rehearsal or review of only one-half of the items. Table 1 schematizes these conditions.

Thus, for each "complete-coverage" review or rehearsal treatment, there were two corresponding partial conditions that utilized the same instructional procedure but covered a shorter set of items in the review or rehearsal sessions. Partial versions were constructed by dividing the six or eight letters in each block of the film into two sets, designated "1" and "2." Thus, the partial versions bearing the designation 1 (A'1, B'1, C'1, etc.) contained one set of 13 items in their review sections, while the number 2 partial versions (A'2, B'2, C'2, etc.) contained the other 13 items in their review sections.

Description of Experimental Versions

The "overt-oral" treatment involved instructing the subjects to call out the phonetic names when letters appeared paired with question marks in a review

TABLE 1

Experimental Conditions, Their Letter Designations, and Total Number of Flights in Each Treatment*

	Overt Oral		Overt Written		Covert (No Feedback)		Covert Plus Feedback		Passive Review		No Review
COMPLETE	A		B		C		D		E		F
	4		4		4		4		4		8
PARTIAL	A'1	A'2	B'1	B'2	C'1	C'2	D'1	D'2	E'1	E'2	-
	2	2	2	2	2	2	2	2	2	2	-

*A', B', C', D' and E' indicate "partial" versions that reviewed only 13 of the phonetic equivalents (see text); A, B, C, D, and E (without primes) indicate versions given review of all 26 phonetic equivalents.

sequence on the screen. The "overt written" version requested the subjects to write the phonetic names on an appropriate practice blank. In covert practice the film narrator directed the subjects to "think" the phonetic names, and in the "covert-plus-feedback" version this "thinking" of the phonetic names was followed, after an interval, by the correct answer spoken by the narrator. In the passive-review version, the letters appeared on the screen in review sections, but student rehearsal was discouraged by having the narrator quickly call out the correct phonetic equivalent of each letter. "No review" consisted of a straight film-showing without any review sections.

Experimental Procedure

Two replications of the experiment were run, with order of flight assignment to specific treatments randomized within each replication. Within each replication, two flights were assigned to each "complete" treatment and one flight was assigned to each of the pairs of "partial" versions. The no-review treatment included four flights per replication (see Table 1).

A full flight of Sampson Air Force Base basic trainees (approximately 60 airmen) was tested in each experimental session, for a total of 48 flights, involving approximately 2,600 men. An announcement that a test would follow the film was made at the start of every session; then the appropriate film version was shown, and a test of immediate recall covering all 26 items concluded the session. The test consisted of a series of slides, each one showing on the screen a letter of the alphabet and nothing more, with the order of presentation of the letters randomized. The trainees were told to write the phonetic equivalents of the letters on answer sheets, and a period of ten seconds was allowed for each letter. In scoring the answer sheets, two points were awarded for each perfect answer. One point was given for one of a specified list of nearly correct answers, and zero was scored for an incorrect phonetic equivalent.

Results

Range of Performance of Groups with Complete Coverage

Grand mean scores for all treatment groups (A-F) that had complete coverage of the items in active or passive reviews, and the grand mean for no review, are summarized in Table 2. Data for each individual flight tested in the experiment, including mean sub-set scores, mean total scores, and mean AFQT scores, are given in the Appendix.

Overall, the 16 flights given complete coverage of the items in one or another of the active-review treatments attained a mean total score of 39.09, or 75 per cent of the possible total (52 points). Compared to this score, the four passive-review flights obtained a mean of 32.04, or 62 per cent, and the eight flights that had no review received a mean score of 29.34, or 56 per cent of the possible total. Hence, the range between the overall grand mean for "active review" and the grand mean for the no-review treatment was 9.75 points, representing nearly one-fifth of the total possible score.

Comparative Effectiveness of the "Complete" Treatments

In order to evaluate differences among the six instructional procedures (four active-review treatments, one passive-review, and one no-review), an analysis of variance was performed on the mean scores obtained by the 28 flights in the complete condition. Another analysis of variance was performed on the mean scores of the flights in the four active-review treatments, i.e., on 16 of the previously mentioned 28 flights, with the aim of determining the reliability of differences among the four practice procedures. Following this, *t*-tests were performed to pinpoint the significant differences between treatments indicated by the analyses of variance.

The results of the two analyses of variance are summarized in Tables 3-A and 3-B. Among the six treatments included in the first analysis (Table 3-A), the overall treatment differences proved to be highly reliable. This outcome parallels the findings of earlier experiments, which have shown with various subject matters

TABLE 2

Grand Mean Total Scores of Groups Given Complete Coverage of the Items in Active or Passive Reviews Compared to the Grand Mean for the No-Review Treatment

Treatment	Number of Flights	Grand Mean
1. Active Review (Complete)		
(A) Overt oral	4	38.59
(B) Overt written	4	39.67
(C) Covert	4	37.29
(D) Covert plus feedback	4	40.81
Overall Active Review (Complete)	16	39.09
2. Passive Review (E)	4	32.04
3. No Review (F)	8	29.34

and student populations that real differences can be obtained in the instructional effectiveness of active review, passive review, and no review. The second analysis (Table 3-B), covering the flights included in the four different active-review treatments, did not provide support for concluding that the four active-review procedures differed among themselves in efficiency.

TABLE 3

A Analysis of Variance for all Six Treatments—Active Review, Passive Review, and No Review—Complete Condition Only

Source	df	Mean Square	F
Treatments	5	116.74	23.6**
Within Groups	22	4.95	--

**P <.001

B Analysis of Variance Among the Four "Active Review" Treatments—Complete Condition Only

Source	df	Mean Square	F
Treatments	3	9.09	1.50*
Within Groups	12	6.05	

*P >.2

Using total scores of the flights included in the first analysis of variance, several t-tests were computed, which are summarized in Table 4. It can be noted,

TABLE 4

Tests of Differences Among Active-Review, Passive-Review, and No-Review Treatments—Complete-Coverage Groups Only

	Comparison	Mean dif.	t	df	p (1 tail)
Combined Group Comparisons	"Passive" minus no review	2.7	2.3	10	<.03
	Combined "Active" minus "Passive"	7.1	4.8	18	<.0001
	Combined "Active" minus no review	9.7	10.0	22	<.00001
Separate Treatments Versus No-Review	Overt oral minus no review	9.2	10.9	10	<.00001
	Overt written minus no review	10.3	10.8	10	<.00001
	Covert minus no review	8.0	7.1	10	<.0001
	Covert plus feed-back minus no review	11.5	11.2	10	<.00001
Separate Treatments Versus Passive Review	Overt oral minus "Passive"	6.6	4.0	6	<.005
	Overt written minus "Passive"	7.6	4.3	6	<.005
	Covert minus "Passive"	5.3	2.7	6	<.02
	Covert plus feed-back minus "Passive"	8.8	4.7	6	<.002

for example, that the mean difference of 2.7, favoring passive review over no review, generated a t of 2.3 (p<.03). Also, the passive-review treatment mean of 32.0 versus the four active review groups, taken together, with a mean of 39.1, produced a t of 4.8 (p<.0001); and the no-review treatment, compared with the same grouping of the four rehearsal treatments, gave a t of 10.0 (p<.00001).

The tests also indicate that the obtained differences favoring the various rehearsal treatments individually over the passive-review and no-review treatments are highly reliable.

Comparative Effectiveness of Partial and Complete Coverage of the Items

The preceding analyses included only the groups given the complete list of instructional items in the active- or passive-review sessions. The following analysis compares the scores of the no-review group and of the complete-coverage groups with the scores of the "partial" groups. These scores are summarized in Table 5.

TABLE 5

Mean Sub-Set and Mean Total Scores of Partial and Complete Groups in Active- and Passive-Review Treatments, and Scores of the No-Review Treatment—Underlining Indicates Coverage of the Items in the Reviews

Treatment	Mean of Sub-Set 1	Mean of Sub-Set 2	Mean Total Score
Active Review			
Complete (A, B, C, & D)	19.06	20.03	39.09
Partial 1 (A'1, B'1, C'1, & D'1)	18.95	16.39	35.34
Partial 2 (A'2, B'2, C'2, & D'2)	14.21	19.24	33.45
Passive Review			
Complete (E)	15.22	16.82	32.04
Partial 1 (E1)	13.37	13.75	27.12
Partial 2 (E2)	12.65	15.99	28.64
No Review (F)	13.61	15.73	29.34

Tests employing mean total scores showed that the superiority of active-review over no-review procedures was highly reliable (Table 4), and this superiority holds also for the sub-set mean scores using those items rehearsed by the partial coverage groups (see Table 5). A t-test between the mean obtained on item sub-set 1 (mean 1) by the no-review flights and the mean 1 for the combined active-review partial-1 flights (A'1, B'1, C'1, and D'1) on the reviewed half of the items showed a difference of 5.3 in favor of the partial flights, $t = 9.0$, $p < .00001$. The parallel comparison for mean 2 (also no review versus combined active review) gave a difference of 3.5 in the same direction, $t = 6.6$, $p < .0001$. These tests indicate that a real gain was found on the items included in the partial-rehearsal treatments when these are compared with no-review treatments.

Still using sub-set means for items rehearsed by the partial groups, additional t-tests were computed against the sub-set means of flights given rehearsal of all the items (complete coverage). In the case of mean 1, a difference of .11 point was obtained, and this is not reliable. With mean 2, a mean difference of .79 point is also not reliable. Hence, there is no basis for believing that the effect of partial coverage is different from the effect of complete coverage for the particular set of items covered in the partial procedure.

The final set of comparisons was designed to assess the differences obtained on sub-set means of active partial review versus no review for scores on items not covered in the partial procedures. The results of these comparisons should reflect the effect on learning additional items of student rehearsal of a portion of the items. These tests are summarized in Table 6.

TABLE 6

Tests of Differences Obtained for Sub-Set Scores Not Covered in Partial Rehearsal, for Combined Active Versus Passive Review, and Versus the No-Review Scores

Comparison	Mean dif.	t	df	p (1 tail)
Mean 1 (not covered) of partial coverage groups (A'2, B'2, C'2, & D'2) minus no review	.6	1.0	14	>.1
Mean 2 (not covered) of partial coverage groups (A'1, B'1, C'1, & D'1) minus no review	.7	1.3	14	>.1
Mean 1 (not covered) of partial coverage groups (A'2, B'2, C'2, & D'2) minus "passive" review	1.6	1.1	8	>.1
Mean 2 (not covered of partial coverage groups (A'1, B'1, C'1, & D'1) minus "passive" review	2.6	2.6	8	<.02

In Table 6 we see that, for mean 1, the opposite partial flights (those that did not review sub-set 1 in the active reviews) differed from the no-review flights by only 0.6 ($t = 1.0$, $p > .1$) and that a similar result was obtained with mean 2 (no review versus A'1, B'1, C'1, and D'1), where $t = 1.3$, $p > .1$. Thus Table 6 shows no evidence for any reliable superiority of partial-coverage active-review treatments over no review on the non-covered items.

An alternative evaluation of the effects of partial active review on non-covered material involves testing the differences obtained between passive and active partial review procedures, also shown in Table 6. In these tests the scores for item sub-sets not included in active reviews are compared to scores for item sub-sets not included in passive reviews. With mean 1, the opposite partial flights having active review did not score reliably better than the opposite partial flights having passive review ($t = 1.1$, $p > .1$). With mean 2, however, the difference of 2.6 favoring active review produced a t of 2.6, p .02.

Thus only one of the four possible tests of differences between partial active review and the control treatments for items not included in the active reviews shows a significant superiority for the partial active review treatments. The other three tests do not show such a difference. This mixed outcome makes a summary statement hazardous. It is interesting to note, however, that of the seven treatment groups, or sub-groups shown in Table 5, only one (the passive-review, partial-1 treatment)—which was the sub-group involved in the lone significant test of the four in Table 6—has a higher obtained score (though only slightly higher) on the unreviewed items than on the reviewed items. Also, this treatment achieved the lowest mean score of all seven treatments on sub-set 2.

Additional Analyses of Results

Prior to computing the analyses reported in preceding sections of this report, a separate evaluation of the results was also made. A scattergram showing test score and AFQT for individual trainees had shown a correlation near +.50. It therefore seemed possible that the precision of the analysis would be increased by use of an analysis of co-variance technique. Hence, flight means on the phonetic alphabet test were adjusted to make allowance for chance differences in the aptitude (AFQT) of the men in the different flights. The tests summarized below were based on such adjusted scores.

Treatment Differences. The first analysis of co-variance was designed to evaluate the reliability of differences among the six instructional treatments. One flight was selected at random from each pair of flights assigned to the complete-coverage conditions in each replication of the design. Adjusted means for these flights grouped by treatments are shown in Supplementary Table 1-A. Supplementary Table 1-B summarizes the results of this first co-variance analysis. Since an F of 7.64 was obtained for variation attributable to instructional treatments, it may be inferred that the differences between mean scores obtained by the different review and no-review groups are real and not due to sampling fluctuation. This inference replicates the outcome of the tests previously reported, when all six treatments were included in the computation. The treatments by interaction of replications also appears to be reliable, and, from the finding that this interaction variance was more than three times as large as the variance within groups, it is inferred that flights were more homogeneous in composition than were random samples of students from all flights.

Treatment and Coverage Differences. Two additional analyses of co-variance were based on data from 30 of the 36 sessions omitted from the preceding treatment differences analysis. Sub-set mean scores were used instead of total scores so that the effectiveness of partial coverage of the instructional items in active and passive reviews could be evaluated for both covered and omitted items. The 30 testing sessions (flights) include all the groups omitted from the previous co-variance analysis except the no-review flights; they comprise six flights from each of five review treatments, or ten flights from each coverage condition (complete, partial 1, and partial 2). Adjusted sub-set grand mean scores by treatments and coverage conditions are shown in Supplementary Table 2.

The first of these two additional co-variance analyses was computed using mean scores on the first item sub-set. Both this and the second analysis, which involved scores on the second sub-set, are essentially tri-dimensional (triple-order interaction) designs, with review treatments, coverage conditions, and experimental design replications constituting the three dimensions. Table 7-A shows the results of the co-variance analysis of first sub-set scores and Table 7-B summarizes the parallel co-variance analysis of second-item sub-set scores.

Both analyses showed significant variance ratios for instructional treatments (four active reviews and passive review) and for coverage conditions (complete, partial 1, and partial 2 coverage). It will be recalled that the analyses presented in Table 3-A (and Supplementary Table 1-B) covered all six treatments, while the analyses reported in Table 3-B, which did not indicate a significant difference among treatments, included only the four active reviews. Hence, these two triple-interaction analyses add the information that, using the adjusted scores, the variance among treatments is significant across the five review procedures, excluding "no review." (The significant F tests for coverage conditions indicate that, for each item sub-set, review versus non-review of the sub-set of items produces a real effect on the scores, as anticipated.)

An overall comparison of learning scores for the different review treatments is shown in Table 8. Mean scores on the two sub-sets of items were summed for each of the six flights under each treatment, disregarding conditions of coverage in the reviews. The resulting mean total scores for the flights were then pooled to get grand mean total scores by treatments.

The grand mean scores from the 30 flights included in the two triple-interaction analyses fall into the same ranking that was obtained in adjusting scores of the 12 flights selected from the complete-coverage condition for the previous co-variance analysis (presented in Supplementary Table 1-B). "Overt-written" practice has the highest learning score, followed by covert rehearsal with feedback, "overt-oral" practice, covert rehearsal, and last, conventional review. The probability of obtaining through chance the same ranking of the five treatments in each of two independent analyses is, of course, slight.

In Table 8 total scores were pooled across coverage conditions to obtain grand means by treatments. This procedure is reversed in Table 9, with treatments

TABLE 7

Analysis of Variance on First and Second Sub-Set Means for Complete and Partial Conditions

A. First Sub-Set

Source	df	Mean Square	F
Treatments	4	946.51	24.45**
Conditions	2	3373.42	406.46**
Replications	1	27.43	--
T x C	8	261.95	--
T x R(1)	4	38.71	--
C x R(2)	2	8.30	--
T x R x C(3)	8	116.62	--
Within Groups	1641	26.70	--

B. Second Sub-Set

Source	df	Mean Square	F
Treatments	4	508.76	6.72*
Conditions	2	1935.03	160.39**
Replications	1	11.61	--
T x C	8	74.37	--
T x R(1)	4	75.74	--
C x R(2)	2	12.06	--
T x R x C(3)	8	50.35	--
Within Groups	1641	23.29	--

*p <.05
**p <.01
(1) Error term for Treatments.
(2) Error term for Conditions.
(3) Error term for Replications.

pooled to show the overall differences among the three conditions of review coverage. It will be noted in Table 9 that the sub-set scores obtained under conditions of complete coverage are quite similar to the sub-set scores of the covered items under partial coverage. The major share of the differences in grand mean total scores for conditions is attributable to the lower sub-set scores that resulted from non-coverage of a sub-set in the review procedures.

Individual Treatment Differences. Using the adjusted sub-set scores of complete-coverage groups (those included in the triple-interaction analyses), *t*-tests were computed between pairs of instructional treatments. In the 20 tests in this

TABLE 8

Comparison of Adjusted Mean Total Scores Pooled over Coverage Conditions as Compared with Adjusted Mean Total Scores of Complete-Coverage Condition

Treatments	Pooled Coverage Condition	Complete Coverage
1. Active Review		
(A) overt oral	35.93	37.73
(B) overt written	37.20	41.06
(C) covert	34.93	35.85
(D) covert plus feedback	36.18	38.83
2. Passive Review (E)	29.76	33.28

TABLE 9

Adjusted Mean Scores Pooled for Five Treatments and Arranged by Conditions of Review Coverage

(Underlining Indicates Coverage of the Items in the Reviews)

Conditions	Mean 1	Mean 2	Total Mean
1. Complete	<u>18.47</u>	<u>19.54</u>	<u>38.01</u>
2. Partial 1	<u>18.03</u>	15.98	34.01
3. Partial 2	13.96	<u>18.61</u>	32.57

series none of the differences among active-review treatments produced probabilities as low as five per cent, but a majority of the comparisons between an active-review and passive-review treatment showed reliable differences. This outcome parallels that of the earlier analysis of raw scores. It will be recalled that the analysis of variance across the four active reviews yielded a *p* of more than .2, while the *t*-tests summarized in Table 4 showed reliable superiorities for each active-review treatment over both passive review and no review.

Discussion

The basic aims of this investigation were: (1) to evaluate the teaching effectiveness of several alternative forms of student rehearsal (commonly referred to as "audience participation"), using a film teaching the phonetic alphabet, and (2) to assess the effects of "partial" rehearsal procedures, i.e., directed practice of one type or another covering only part of the full list of instructional items. The central point here was to determine whether or not active review of some items has an effect upon the learning of additional items. While there is a growing literature of studies on instructional-media learning, supporting the contention that some kind of active recitation or rehearsal is superior to conventional review, there are no earlier studies involving paired-associate learning that provide comparison of the wide range of different rehearsal procedures evaluated in the present study. In addition, previous studies of only partial coverage in rehearsal, instead of the full list of instructional items (e.g., the Chapter-18 Michael and Maccoby experiment using the civil-defense film, "Pattern for Survival") have not provided evidence that is readily generalizable to paired-associate situations such as learning of the phonetic alphabet.

With respect to the first general aim of this experiment—evaluation of the relative teaching effectiveness of different active-review procedures—the results show the same superiority of active review over conventional passive review that has been found in previous studies. No evidence was obtained, however, for marked differences in teaching effectiveness among oral recitation, written practice, "thinking" of answers, and "thinking"-plus-"feedback" used as alternative active-review procedures. Among these four active-review techniques, the single comparison that might have seemed most likely to show a reliable difference is the covert treatment (mean = 37.3) versus the covert-plus-"feedback" treatment (mean = 40.8). In their experiment, Michael and Maccoby had concluded that provision of knowledge of the correct answer (feedback), after the students had formulated their answers in practice sessions, was the most important factor in determining the extent of learning gains from rehearsal. In the present experiment, however, this comparison of covert practice with "feedback" and covert practice without "feedback" showed only rather slight differences, with a t of only 1.83 (single-tail p of .06).

In comparing practice procedures such as oral, written, and covert rehearsal, the experimenter is varying the instructions given to the students. Viewed from the vantage point of the class proctor, student responses to these different instructions are quite different—verbalizing versus writing versus silent practice. The lack of sizable or significant differences among the different practice instructions may suggest, however, that the crucial consideration is not the type of response that the students are instructed to make but rather the fact that the students are instructed to practice, and actually do practice, associating the question-and-answer pairs. Not enough data are available as yet to permit a confident prediction that these findings would apply to other training tasks and situations. If, however, additional research shows covert rehearsal to be as effective as it was found to be in this experiment and in the Michael and Maccoby study, this procedure might find wide application in programmed instruction.

The second general question that this experiment was designed to answer concerns the effectiveness of partial coverage in active-review sessions on the full list of instructional points. No reliable differences were found between partial coverage and complete coverage on the specific set of items that was rehearsed. Further, scores on the items that were not covered in the active-review sessions of the partial-coverage groups were not significantly different from scores of the no-review groups, and differed from scores of the partial-coverage groups in the passive-review treatment for only one of the two sub-set means. Thus, while no evidence was obtained here that partial rehearsal is inferior to complete coverage for the items included in the partial procedure, the results of this study of partial rehearsal with paired-associate learning are quite similar to the results obtained by Michael and Maccoby (Chapter 18) with factual items, i.e., learning gains occurred with rehearsed items, but no marked carry-over to additional items has been established.

Some operational situations may preclude covering all the instructional points in practice sessions. For these situations it would seem most fruitful to select the most important items for practice, while not completely abandoning hope for improvement in the learning of additional points. Even if no direct "motivational effect" of partial practice occurs [at least under conditions of generally good motivation], gains on additional items may result if the practiced and better-learned items are related to learning additional points. One situation of this sort in which carry-over of learning gains would be expected is a task that has, as a central element, learning a method of solution. With a film that teaches an instrument-reading skill, for example, use of several illustrative settings in practice sessions would be expected to promote ability to read additional settings (since illustrative settings should increase learning of the proper method of reading the instrument, as well as learning of the unique problems presented by the particular settings). [It will be recalled that this finding was, in fact, obtained by the study reported by Kimble and Wulff in Chapter 15.]

Furthermore, student rehearsal may increase learning of additional items by interrupting responses that are incompatible with learning. Faison, Rose, and Podell (1955) studied attentiveness in an unfavorable learning situation, where a rather dull and instructionally weak film was shown under unsatisfactory (though, then at least, not particularly unusual) showing conditions. As a result of merely having students stand up and stretch after each main section of a film, both attentiveness and learning were raised, and it seems likely that periodic active-review sessions might have the same tendency to increase attentiveness when the learning situation is unfavorable to normal attentiveness. [See also the results under low-motivation conditions reported in the second part of Chapter 19.]

SUPPLEMENTARY TABLE 1-A

Grand Mean Total Scores of Flights Randomly Selected for the First Analysis of Co-variance

Treatment	Number of flights	Grand Mean*
1. Active Review		
(A) overt oral	2	37.73
(B) overt written	2	41.06
(C) covert	2	35.85
(D) covert plus feedback	2	38.83
Overall Active Review	8	38.37
2. Passive Review (E)	2	33.28
3. No Review (F)	2	28.99

*Each flight mean score was adjusted for AFQT by the equation:
$\overline{X}_{adj} = \overline{X}_{obt} - (.1907)\,(\overline{X}_{AFQT} - \overline{X}_{AFQT})^2$.

SUPPLEMENTARY TABLE 1-B

Analysis of Co-variance for All Six Treatments—Active Review, Passive Review, and No Review—Complete Condition Only

Source	df	Mean Square	F
Treatments	5	2212.00	7.64*
Replications	1	.261	--
Interaction	5	289.64	3.04**
Within Groups (AFQT)	674	95.20	--

* $p < .05$
**$p < .02$

SUPPLEMENTARY TABLE 2

Adjusted Sub-set Mean Scores for 30 Flights Representing Five Review Treatments and Three Conditions of Item Coverage

Treatment	Condition		Number of Flights	Grand Mean 1*	Grand Mean 2**
Overt oral	(A)	Complete	2	19.59	20.30
	(A'1)	Partial 1	2	19.66	17.40
	(A'2)	Partial 2	2	12.40	18.26
Overt written	(B)	Complete	2	18.52	19.53
	(B'1)	Partial 1	2	20.67	17.33
	(B'2)	Partial 2	2	14.79	20.11
Covert	(C)	Complete	2	18.49	19.82
	(C'1)	Partial 1	2	17.95	15.51
	(C'2)	Partial 2	2	14.25	19.05
Covert plus feedback	(D)	Complete	2	20.41	21.03
	(D'1)	Partial 1	2	17.71	15.53
	(D'2)	Partial 2	2	15.06	19.17
Passive Review	(E)	Complete	2	14.99	16.97
	(E'1)	Partial 1	2	14.48	13.88
	(E'2)	Partial 2	2	13.14	16.43

* Each flight mean score was adjusted for AFQT by the equation:
$\bar{X}_{adj} = \bar{X}_{obt} - (.1134)(X_{AFQT} - \bar{X}_{AFQT})^2$.

**In the second sub-set mean adjustment equation, $\beta = .1025$.

CHAPTER 29

THE ROLE OF CLASS-DESCRIPTIVE CUES IN PAIRED-ASSOCIATE LEARNING[1]

J. Jepson Wulff
Lawrence M. Stolurow

Introduction

The findings of several learning studies (e.g., Detambel and Stolurow, 1956; Kurtz, 1955; Osgood, 1953; Spence, 1936; Stolurow, 1956; Wulff, 1954) make it clear that stimulus processes are of basic importance to learning. The present experiment was designed to test the usefulness of one way to conceive of stimulus processes as they are related to cue utilization. The hypothesis is based upon a conception of the stimulus event in the S-R paradigm that differentiates between the stimulus object and the associative stimulus (s). Specifically, it is assumed that the associative stimulus, s, which is said to be associated with the "correct" response in paired-associates learning, has three principal properties: (a) It is an "implicit" response to features of the stimulus object, the explicit or manipulated stimulus; (b) it has stimulus properties; and (c) it is a modifiable event rather than one with a fixed relationship to the properties of the stimulus object. The associative stimulus is said to be modifiable because it is assumed that the specific "implicit" responses that function as the associative stimulus event in a given instance are in part a function of the learning conditions attending the formation of the association. Thus, if the stimulus objects in a paired-associates list afford the learner opportunity to make implicit responses to alternative features of the objects, one set of circumstances surrounding learning may favor one implicit response and another set of circumstances may favor a different implicit response, or even a complex of implicit responces, as an associative stimulus.

[1] This paper is a revision of a report by the same title, which appeared in the Journal of Experimental Psychology, 1957, Vol. 53, 199-206; and which was distributed separately by the Air Force as AFPTRC-TN-57-80, June 1957. The material in this paper was also presented at the twenty-sixth annual meeting of the Midwestern Psychological Association at Columbus, Ohio, in 1954. The assistance of Dr. R. Capobianco is gratefully acknowledged. This research was supported in part by the United States Air Force by Contract No. AF 33(038)-25726 with the University of Illinois, monitored by A. A. Lumsdaine for the Training Aids Research Laboratory of the Air Force Personnel and Training Research Center. Dr. Wulff is now with Psychological Research Associates, and Dr. Stolurow is at the University of Illinois.

In this experiment the usefulness of this way of conceiving of the stimulus event was examined by comparing the level of learning resulting from two different training conditions. According to the above formulation, one of these training conditions would be expected to favor the utilization of combinations of cues, consisting of the cue that was both necessary and sufficient for correct response and an ancillary cue, as contrasted with the other condition, which would be expected to favor the utilization only of an individual cue that was necessary and sufficient for correct response. Furthermore, under the circumstances employed, the utilization of cue combinations of the type described should give rise to faster learning than the use of only an individual cue.

Problem

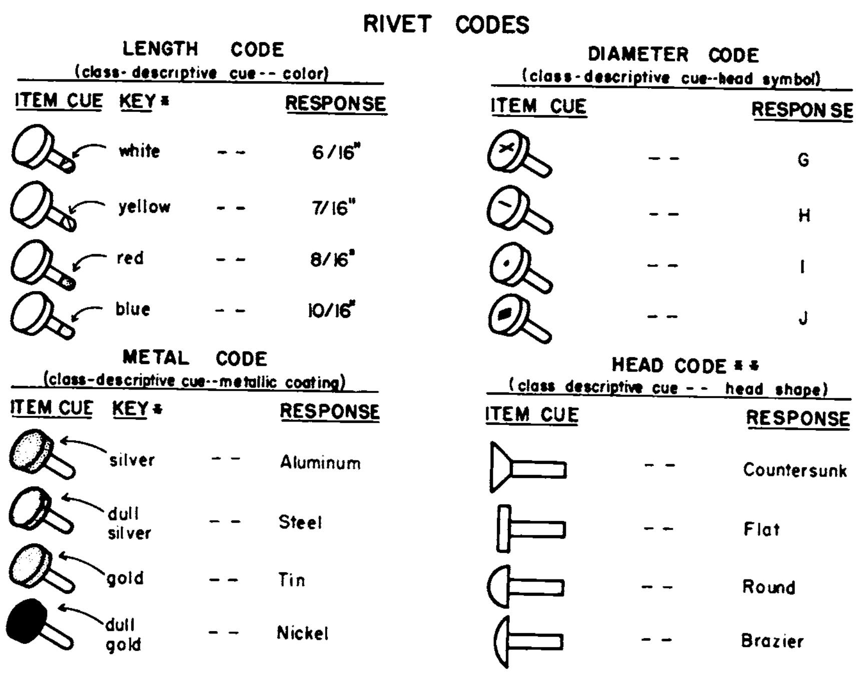
Frequently, a list of paired associates may be subdivided descriptively into several sub-groups by categorizing the stimulus terms on the basis of features that are the same within stimulus-term sub-groups, but different between them. Features of stimulus objects that can be utilized in this way can be called class-descriptive features. For example, a list of paired-associates with colored stimulus terms, each of which differs from all others in shape, may be divisible into sub-groups of paired items on the basis of color. In such a list it is clear that color need not be attended to by the learner, since shape alone will discriminate all of the stimulus terms. Nevertheless, with a long list of such paired associates one might expect that the color of each stimulus term would be attended to in addition to its shape, since this would have the effect of simplifying the task. Thus the long list would become a number of smaller tasks discriminable from each other on the basis of color. Under this condition, one would expect color to be especially helpful to the learner if a pre-differentiated response class corresponded to each color or class-descriptive feature. For example, if the response terms for all of the yellow stimulus terms were numbers, and if no other stimulus term were paired with a number response, then yellow might be a useful class-descriptive feature of the stimulus terms to denote the "number" class of responses. However, in this case, since yellow could not be used alone as a cue for a specific response to any particular stimulus term, it would not be a necessary cue in learning to make correct responses to individual stimulus terms.

It is hypothesized that a class-descriptive feature of the stimulus terms, such as color in this example, may give rise to an implicit response that acts with the implicit response to each item-descriptive feature (a shape, for example) as a compound stimulus event that is associated with the correct explicit response. If the implicit response to a class-descriptive feature of the stimulus terms has become associated, as a stimulus, with its response class, it can be expected to increase the probability of a correct response and decrease the probability of competing responses from other response classes through bonds between the associative stimulus deriving from the class cue and the terms of the correct response class. It follows from this reasoning that learning conditions that favor the utilization of class cues (as well as item cues), as compared with those conditions which do not favor the use of class cues, should facilitate learning the total task. The two methods of training employed in the present study were designed to permit a test of this hypothesis; one method favors the use of class cues and the other does not. It should be noted that the term, "cue," is used here to denote a feature of the stimulus object that gives rise to an associative-stimulus event.

Method

Learning Task

All S's used in the experiment were required to learn to interpret a specially devised code for rivets. As shown in Figure 1, four different features of rivets were coded: length, diameter, type of metal, and type of head. Four different values of each of these four features were included in the task, making a total of sixteen paired associates in which the codes were treated as stimulus terms, and in which the values or names assigned to the codes were treated as response terms. Subsequent to experimental training by one of the two training methods used, S's were required to decode and encode rivets in order to measure the amount of learning accomplished under the training methods used. One of these methods was designed to foster the use of class-descriptive cues; the other was not.



* While words are used in the present figure to label the differently marked features of these drawings, the actual colors signified by these words were used in the training and Test I materials of the experiment.

** Head type was not coded in the conventional sense of the word, "code," since the shape of the side view of the manufactured head was employed as the cue; no abstract symbol was substituted for it as was done in the case of the other codes.

Figure 1. Summary of the rivet code used in the experiment.

Training Materials

Inspection of Figure 1 will reveal that the codes were designed in such a way that both pre-differentiated class-descriptive and item-descriminative cues are available for both stimulus and response terms. For example, it can be noted in Figure 1 that "color on tip of rivet," is a class-descriptive cue for one sub-set of four stimulus terms, and that on the response side "length" responses or "sixteenths-of-an-inch" responses are appropriate for this class cue. Within this class of stimulus terms, the specific colors—white, yellow, red, and blue—provide item-descriptive cues. Each color denotes a specific length response and is itself a necessary and sufficient cue for that response. The three other sub-sets of stimulus terms can be defined by the following class-descriptive cues: black mark on rivet head, metallic color of rivet head, and shape of side view of rivet head. Each of these three cues denotes a different response class (i.e., diameter, metal, and head type, respectively), and within each of these sub-sets of stimulus terms there are four item-descriptive cues, each of which is uniquely related to a different response item. All item- and class-descriptive cues were assumed to be minimally confusable to the S's.

Experimental Conditions

In order to implement the experimental comparison, this code was taught by two different methods of presentation: a class-organized method and an object-organized method.

Class Organization of Materials. This method of presentation was designed to provide conditions that would foster the use of class-descriptive cues. To this end the four paired associates for the length code (e.g., white—6/16 inch, yellow—7/16 inch, red—8/16 inch, and blue—10/16 inch) were all presented together, and the same procedure was used for diameter, metal, and head type. Thus there were four 5x8-inch training cards, each with a different class of four paired associates. An example of a card constructed in this way is shown in Figure 2. There is a certain letter name for every different rivet diameter. Letters like A, B, and C are used for small diameters; letters like X, Y, and Z are used for rivets with a large diameter. On the top of every rivet there is a black code mark which tells the diameter of the rivet. In this task you will learn the code for four of the most commonly used diameters: G, H, I, and J.

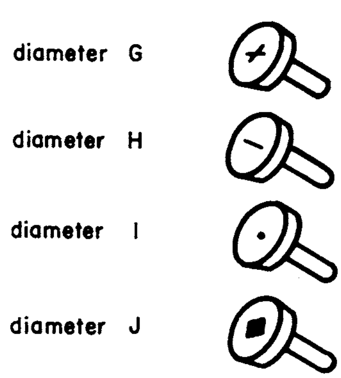
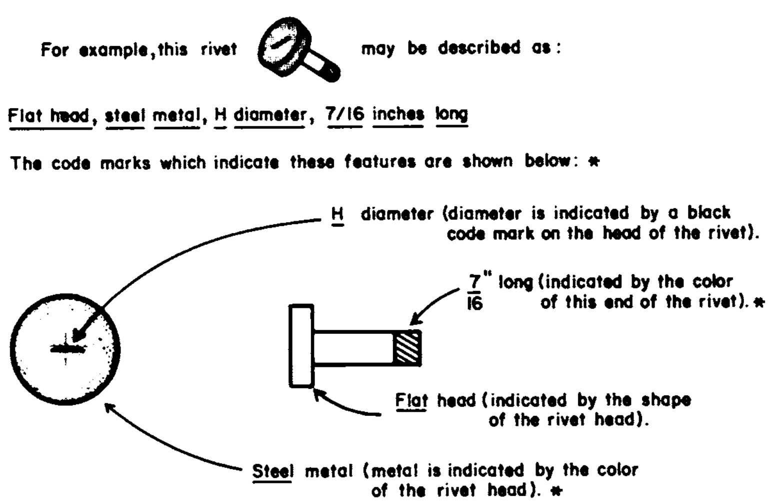


Figure 2. A training card illustrating the class-organization method of presenting codes. This figure shows the presentation of the code for diameter.

Object Organization of Materials. This method of presentation was designed to provide conditions that would not foster the use of class-descriptive cues. In this method each of four training cards displayed a different picture of a sample rivet. Every picture showed one cue from each of the four code classes. With this method the information about the codes is reorganized into sub-tasks about a sample object, rather than about a rivet-feature class. The materials for object-organization training were prepared by selecting four sample rivets, each with four different code examples so that all sixteen code items were used. The coded features of each rivet picture were labeled in the manner indicated in Figure 3. It can be seen that the balanced condition—four code sub-tasks, each of which contained four cue-response items—made it possible to equate the class- and object-organized training with respect to number of items presented at one time.

The first of these two arrangements—class organization—presumably favors the association of each class of responses with a separate class-descriptive cue. Thus, when class organization is used, the association of the class cue, "color on the tips," can readily be associated with all of the responses in the length class—6/16, 7/16, 8/16, and 10/16 inch. In contrast, with object organization of the training materials this type of association can take place only if the learner can recall (or anticipate) all the response terms for each class as each sample rivet is presented.



*The training materials used in the experiment contained actual colors where shading is used in this figure; i.e., the head was dull silver and the tip was yellow.

Figure 3. A training card illustrating the object-organization method of presenting codes.

Training Cards. The training cards for each method of presentation were arranged into a pack of 10 cards. First, there were two introductory cards, then four cards explaining the four codes (as shown in Figures 2 and 3), and finally there were four review cards. The review cards did not include the verbal heading shown for the two cards in Figures 2 and 3; otherwise they were the same. The two introductory cards were exactly the same for both methods of training. The first contained instructions to the trainee for the use of the cards; the second explained what was meant by the length of a rivet, by the diameter of a rivet, and by the metal and head-type of a rivet.

Test Materials

Two tests were used. One required the S's to interpret the codes applied to sample rivets (decoding); the other required them to tell how rivets with given characteristics would be coded (encoding).

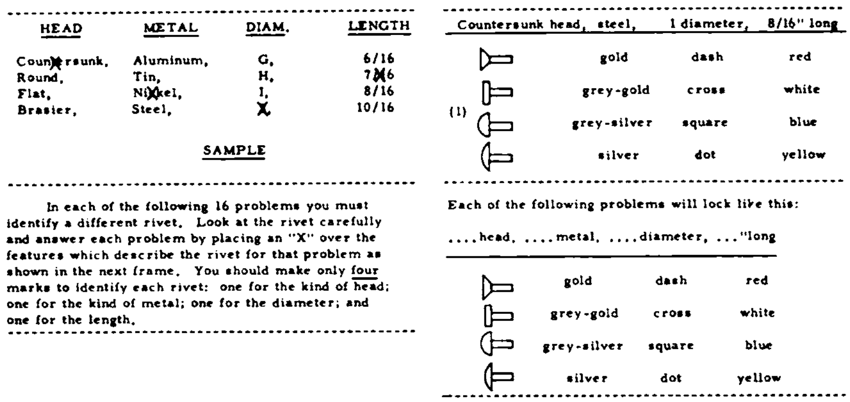
Test 1—Decoding Rivets. The first test required S's to interpret 16 rivets coded in such a way that each cue was used four times but no pattern of three cues was represented more than once. To each of 16 rivets S's made four responses.

As an experimental device, only two different actual lengths and two different actual diameters of rivets were used. Thus the shorter-length rivets were coded to represent either 6/16 or 7/16 inch, and rivets with the longer length were coded to represent 8/16 or 10/16 inch. Similarly, smaller-diameter rivets were coded to represent either I or J, etc. This was done to force S's to use the code rather than the perceived length or diameter of the rivet to determine their responses in describing the rivets. To eliminate weight cues all the rivets were made of aluminum.

Each of the 16 test rivets was placed in one of the 16 numbered cells of a tray. Answer scrolls that fit into a reading-rate recorder—a machine that allowed S's to expose in a window one answer panel at a time—were prepared with 16 numbered blocks or panels. Each panel contained all 16 of the possible response items grouped according to the four sub-tasks; i.e., by head, diameter, metal, and length. S's were instructed to pick up each rivet in turn and to cross out the four correct responses in the answer panel. A sample answer panel is shown in Figure 4.

Test 2—Encoding. For the second test, eight more answer panels were prepared on the same scroll. At the top of each of these panels there was a description of a rivet in terms of the response item (for example, "Flat head, aluminum, G diameter 6/16" long"), and below the rivet description all 16 code-cue stimuli were listed. They were arranged in four arrays as in Test 1 (see Figure 4), and S's crossed out the four appropriate code descriptions for each rivet. S's made 32 responses in this test, four responses to each of eight rivet descriptions. Each of the 16 response names was used twice in this test, and no two rivet descriptions had as many as three items in common. The printed instructions for Test 2 (encoding) are presented in Figure 4.

The S's paced themselves in taking both of the tests; there was no time limit.



Test 1--Decoding   Test 2--Encoding

Figure 4. Instructions and a sample item for Test 1 and for Test 2, both of which were printed on a scroll. Only material between dashed lines was exposed at any one time. In Test 1, S's decoded 16 actual rivets displayed in a tray; in Test 2, S's indicated, also by crossing out one item in each answer column, how to encode eight different rivets.

Procedure

The order of events for each S was as follows: (a) read each of the 10 training cards once; (b) take Test 1; (c) take Test 2.

Before training began, S's were instructed to read the cards one at a time at their normal rate, not turning back to cards they had already finished. The total time spent in reading the training cards was measured by an ink pen-recording mechanism built into the reading-rate recorder. The machine's recording system moved a pen horizontally across a scroll at a constant rate while S read a card.

Testing was begun as soon as the training cards had been finished, and no break was allowed between tests.

Design

Each of the two training methods was used to train a different group of 20 S's. Half the S's trained by each method were Air Force trainees with high mechanical-aptitude scores (high-ability group) and half were trainees with low mechanical-aptitude (low-ability group) scores as measured by the mechanical-aptitude index of the Airman Classification Test Battery (Daily, 1948). Two S's were trained at a time. With two treatments this meant that four S's were necessary to complete one block (replication) in order to equate the frequency within a block with which the two reading machines appeared. All S's within an experimental-design block had the same mechanical-aptitude index; two of the four were assigned to each treatment in an order determined by reference to a table of random numbers. Since S's were used as they were available, this procedure resulted in an order of blocks that was not sequentially biased with respect to ability level. That is, all high-ability S's were not run either first or last, but rather high- and low-ability blocks were intermixed. The design was therefore a two-by-two factorial (Training Method x Mechanical Aptitude Index) with 10 S's in each of the four cells.

Results

The principal results in terms of mean number of correct responses in Test 1 and Test 2 are summarized in Table 1, by training method. It can be seen that S's given class-organization training performed reliably better than those given object-organization training. The overall difference in performance between these two

TABLE 1

Mean Number of Correct Responses for Test 1 and Test 2 by Experimental Conditions

Test	Variable	Group	Type of Organization of Materials: Class Organization (N = 20)	Type of Organization of Materials: Object Organization (N = 20)	Reliability of Differences
Test 1 (64 items)		Overall	48.0	38.35	P <.01
	Aptitude	High	54.9	46.2	P = .05
		Low	41.1	30.5	.05 >P >.02
	Sub-task (16 items)	Head	13.95	13.70	*
		Metal	9.30	7.20	*
		Diameter	12.20	9.25	*
		Length	12.55	8.20	*
Test 2 (32 items)		Overall	24.65	19.80	P <.05
	Aptitude	High	27.3	24.8	P <.05
		Low	22.0	14.8	

* P values are not appropriate for these comparisons since the differences are not independent of the values presented above.

groups on Test 1 is 9.65 correct responses, which is reliable at the .01 level, as assessed by an analysis of variance (Table 2). Similarly, the difference between these two training conditions on Test 2 of 4.85 is reliable ($P < .05$).

As indicated above, high and low mechanical-aptitude groups were included in the design. This dichotomy was included to increase the precision of the experimental design rather than because of any primary interest in performance as related to aptitude level. Nevertheless, the aptitude groups were compared separately as shown in Table 1, and the performance difference between training conditions is found to be reliable for each of the aptitude groups.

TABLE 2

Summary of Analyses of Variance

Source	df	Test 1		Test 2	
		MS	F	MS	F
Training Method (M)	1	931.22	7.68**	235.22	6.40*
Aptitude (A)	1	2,175.62	17.96**	585.22	15.94**
A x M	1	9.02		55.22	1.50
Within	36	121.10		36.70	

*P <.03
**P <.01

The mean number of correct responses per sub-task is also shown for Test 1 in Table 1, by training method. It can be seen from the tabulated means that the superior performance of students trained by the class-organization method is consistent for all sub-tasks, indicating that the overall comparison does not reflect merely a large superiority with respect to one or two sub-tasks.

The difference in mean speed of reading the training materials is small and unreliable for the two training groups (class training, 320.49 sec; object training, 357.66 sec). This finding makes the interpretation of the test results clear. Since the two groups spent about the same amount of time studying the same content differently organized, it is apparent that organization of the training materials, rather than reading time, is related to the differences in test performance.

Discussion

The results clearly demonstrate that with these experimental materials class-organized training materials are more effective than object-organized materials. This outcome is consistent with the expectation that class-organized training conditions would favor the development of associative-stimulus events involving implicit responses to class cues, which in turn would foster learning the total task. Thus one may interpret these results in terms of the associative-stimulus concept stated above in which it is assumed that the stimulus event, $\underline{s}$, is modifiable.

If one chooses to interpret the present findings in terms of class and item cues, one must recognize that the role of the class-descriptive cue can be formulated in alternative ways that are compatible with these data. As it was stated, the hypothesis tested in this experiment does not provide a basis for differentiating between two alternative paradigms that could be postulated to account for the learning process involved.

One of these paradigms would assume a dual habit structure. For example, an associative stimulus ($\underline{s}^c$) based on a class cue may become associated with all the response terms ($R^c\underline{s}$) for that class as though the list of response terms were a single response. An associative paradigm for this habit can be represented for Class 1 as indicated below.

$$S^{c-1} - \begin{cases} R^{c-1}_{i-1} \\ R^{c-1}_{i-2} \\ R^{c-1}_{i-3} \\ R^{c-1}_{i-4} \end{cases}$$

(First habit in a hypothetical dual habit paradigm)

The second habit would be a simple association between the correct response term and the implicit response to the item cue (e.g., for item one, $\underline{s}_{i-1}$-R_{i-1}; for item two, $\underline{s}_{i-2}$-R_{i-2}, etc.). In the convention adopted to represent the habit elements, the superscript $^{c-1}$ denotes Class 1 and the subscript $_{i-1}$ denotes Item 1. In this experiment, Class 1 might be "color on rivet tip" and the associated responses for the first habit "lengths 6/16, 7/16, 8/16, and 10/16 in." Length as a response class is represented as R^{c-1}, and each item response in terms of 16ths of an inch is represented by R_{i-1}, R_{i-2}, R_{i-3}, and R_{i-4}. For the second type of habit an $\underline{s}_{i-1}$ event might be "white" and its response "6/16 inch." As indicated with this assumed type of associational structure, item cues would be responded to as a separate step in performance following the response to the class-descriptive cue.

The role of class-descriptive cues in the dual habit paradigm may be contrasted with a process in which a single implicit response to each combination of class and item cues is acquired. This state of affairs can be represented by the following associative paradigm:

$$\underline{s}^{c-1}_{i-1} \quad - \quad R^{c-1}_{i-1}\,.$$

From the point of view of this second paradigm, the stimulus event is a single unique compound of class- and item-descriptive cues. The stimulus event for the same example as used above would be "white color on rivet tip" ($\underline{s}^{c-1}_{i-1}$) and the associated response "length, 6/16 inch" (R^{c-1}_{i-1}).

If the latter condition existed, one would expect positive transfer of stimulus learning in an A-B, A-K transfer situation involving class-descriptive cues. Positive transfer of this type has been found by Wulff (1954) and by Kurtz (1955). Wulff's experiments, however, did not involve response sub-classes. In Kurtz's experiments, the item-descriptive cues were obscure, which encouraged the learner to attend first to class cues; and one might hypothesize that, with obscure item cues, class training would be facilitating for a different reason from that assumed in the present study. Thus, in Kurtz's study the class conditions would force the learner to attend to item cues early in learning. This explanation does not apply to the present training situation. The available data do not permit a decision regarding which of the two types of habit structure the class training induced.

Summary

It was postulated that the stimulus unit represented by S in the associative paradigm can be usefully described as: (a) an "implicit" response to features of the stimulus object; (b) an event with stimulus properties; (c) a modifiable event rather than one fixed by the stimulus object. It was hypothesized that when the stimulus objects in a list of objects have class-descriptive features, these features may give rise to implicit responses that become a part of the associative-stimulus event. In this case the associative-stimulus event is said to be compounded of implicit responses to class-descriptive cues and item-descriptive cues.

The utility of this conception of the stimulus unit was tested by comparing the rate of paired-associates learning for a given task under two different conditions of training. The paired-associates task was one for which learning was expected to be facilitated if combinations of implicit responses to class and item cues functioned as the stimulus events. Two training conditions were selected so that one would favor the formation of such compound stimulus events; the other would not. The training condition that favored the use of combinations of cues resulted in a reliably high level of performance with a given amount of training.

CHAPTER 30

THE RELATIONSHIP BETWEEN "WHAT IS LEARNED" AND "HOW IT'S TAUGHT"[1]

J. Jepson Wulff
David L. Emeson

Often it is useful to discriminate between two kinds of human-performance research—learning research and training research. In a learning-research study the objective is to examine the learning process as it is going on. For this reason conditions are deliberately chosen so that learning is distributed over several trials. Then error behavior, for example, may be examined as learning progresses slowly. In training research, the objective is quite different; here the objective is most often to discover training conditions that will facilitate learning and, in the ideal case, will bring about adequate learning after only one practice trial. Both of these kinds of research are, of course, necessary, and the two are related to each other. However, since the immediate objectives of these two kinds of research are different, it is important to discriminate between them.

This chapter reports a training-research study. Its primary objective is to demonstrate the utility of one important step in the development of effective training materials.

Ordinarily the deliberate and careful development of training materials is justified only for those practical training problems that involve some degree of difficulty. If the end performance desired can be obtained by a readily improvised training procedure, there will be so little room for improvement that the effort required to plan effective training materials need not be expended. When the development of carefully planned training materials is justified, however, the development of the materials should be guided by two major principles.

---

[1]This study was conducted at the Maintenance Laboratory of the Air Force Personnel Training and Research Center. This report is based upon a paper presented at the annual meeting of the Rocky Mountain Branch of the American Psychological Association, 11 June 1956, by David Emeson. It was also reproduced as a multilithed AFPTRC Technical Memorandum (Emeson and Wulff, 1957) but was given only very limited distribution. A detailed abstract was published by Lumsdaine and Glaser (1960, pp. 616-618). D. L. Emeson is now on the staff of The Children's Memorial Hospital, Chicago. (See also Footnote 1 to Chapter 29.)

The first of these principles is that the training materials should provide opportunity for the learner to make correct responses to the cues he must use to perform correctly in the criterion situation. The second principle is that the training materials should provide training conditions over and above those required to provide the opportunity to respond correctly, and these training conditions should foster making correct responses to relevant cues. Both of these principles must be observed if the planned training materials are to be most effective.

It is often easy to provide opportunity for correct responses in a training situation. For example, opportunity for responding correctly can be provided in typewriting training simply by providing the learner with a typewriter. In the typewriting training situation it is also relatively simple to satisfy the second part of the first principle—opportunity to respond during training to those cues upon which correct responding in the criterion situation must depend. To provide for the opportunity to use these job cues during typing training, it is only necessary to provide the learner with copy similar to that which he must use on the job.[2] However, if we were to place the learner alone in a room with appropriate stimulus materials—a typewriter and nothing else—and then require that he learn to type, we would expect that the learning process would not be very efficient, even though all of the conditions of the first principle given above were met. More must be done to provide for effective training, and that is the reason for the second principle.

The second principle can be explained most conveniently by choosing another learning task as a basis for the discussion. One that will do well for this purpose is the familiar task of learning to receive, decode, and transcribe information presented aurally by Morse code at a high rate of speed. All that is required to provide a learner the opportunity to make correct responses to relevant cues for this task is to present him with high-speed Morse-code signals and give him a typewriter or pencil and paper. Given these stimulus materials and tools needed for responding, nothing else is required. However, most instructors would agree that providing this kind of opportunity would not be sufficient to foster learning to perform the task; some other features must be provided in the learning situation. All features of the learning situation over and above those providing opportunity to respond to relevant cues intended to foster learning, may be called training conditions. Thus, if we start training by presenting the Morse-code signals at a slow rate of speed, the lower rate of presentation is a training condition. If we intersperse practice trials during which an instructor tells students what responses are correct, this instructor activity is a training condition. If we space the trials during which the learner is told the correct responses according to a schedule, that schedule is a training condition.

---

[2] The use of improper cues in typing training might be exemplified by the use of individual letters presented over earphones. If these were the only stimulus materials used, the learner would never practice responding correctly to cues that he would have to use in typing from manuscript, and transfer to job performance could be expected to be relatively poor at first if job performance required copying from manuscript.

Quite obviously, some of the training conditions that might be superimposed on the opportunity to respond to relevant cues for the Morse-code task will be more effective than other conditions in terms of the training time required to achieve criterion performance levels. Therefore, a major task confronting the training expert is the selection of the best possible training conditions for the task to be learned. This paper is concerned with the description of one strategy for selecting effective training conditions.

One way to devise training conditions is to ask, "What training conditions have research and experience demonstrated to be effective in all learning situations?" In answer to this question we will find that knowledge of results is advocated, that spaced practice is generally better than massed practice, that active participation by the learner is better than passive listening, and so on. These answers and the training conditions we must set up to implement them do not, however, usually require us to examine the specific task to be learned. The training conditions called for are, by and large, good for any learning situation and do not require us to ask, "What must be learned?" Knowledge of results, active participation, and the like, are prescriptions for training that are similar to bed rest and a bland diet as prescriptions for sick people. That is, rest and diet are generally good for sick people, and we need not diagnose the disease in order to recommend them. However, this kind of prescription for a sick person may sometimes be about as effective as a prescription for training conditions that ignores the specific performance to be learned. We might be able to prescribe much more effective training conditions if we were to concern ourselves with identifying what it is that must be learned.

It would be unfair to create the impression that all training research has been concerned with the identification of training conditions that (hopefully) are effective for all learning problems. Some training research has also been devoted to the development of methods for categorizing tasks in ways that are claimed to be useful for determining the training methods to be employed (e.g., Cotterman, 1959); but in general these efforts miss one major point which this study is designed to exemplify. That is, these kinds of efforts have been directed toward categorizing the criterion performances rather than toward categorizing the kinds of learning activity required for the learner as he is learning the criterion performance. This is certainly an important distinction, because we employ training conditions to foster learning, and what activity must be learned (by engaging in it during training) is often not very well defined by categorizing the criterion performance. The actual learning activities required for the learner are a function of the difference between his initial capability when he enters the learning situation and the criterion performance to be required of him before he is through. They are not a simple function of the criterion performance alone.

The importance of this difference can be exemplified by again using a Morse-code learning example, where the incoming code signals are to be typed by the trainee. No matter what the initial repertoire of the learner may be, the criterion performance in this task remains constant. However, the learning problem will change from learner to learner if the learners differ with respect to initial repertoires. Thus, the learning problem for one learner whose initial capability does not include typing skill is different from that of another learner who can type even before training begins. What must be fostered in a training situation is not simply the criterion performance; rather it is the process of going from some initial repertoire to the final performance that must be fostered by training conditions. This is a process of change, the nature of the change being dependent upon the difference between the initial repertoire and the final performance.

When one attempts to categorize a learning problem by reference to the criterion performance alone, one does not take account of the fact that learning is a process of change. The implication of these observations is that if we wish to prescribe training conditions that will be most appropriate for a given learner, we must identify the kinds of learning activities that he must engage in during the process of change.

The strategy exemplified by the demonstration reported below, which is the major subject of this paper, is based on this notion. That is, the strategy is to identify the kind of learning problems posed by the difference between the criterion task and the initial repertoire of the learner, and then to select training conditions appropriate for the identified learning problems. Thus the approach is based on the idea that the learning required in the mastery of most practical tasks involves several different kinds of learning activities, and that different learning processes or activities require different training conditions. We question the usefulness of an approach to training research that is focused upon training methods that would be effective in all learning situations, or an approach focused on training methods that are good for a specific kind of criterion-performance category without reference to the kinds of learning processes actually to be involved. It is our expectation that effective training conditions can be matched to learning processes, but not to learning in general or to criterion-task categories. In order to develop good formalized training procedures, the question, "What is learned?" must be answered by an analysis of the difference between the criterion task and the initial repertoire of learners to reveal the kinds of critical learning processes involved. This analysis must also identify the order in which the learning activities must be fostered for efficient learning. "How it is taught" must then be determined by selecting the kinds of training conditions and the order in which they will be used to foster the critical learning activities identified in the analysis.

The objective of this chapter can now be restated more specifically. The main objective is to demonstrate that effective training materials can be developed by employing a two-step strategy. The strategy is to analyze a training task for the purpose of determining the kinds of learning problems embedded in it, and then to develop training conditions designed to foster the kinds of learning uncovered by the analysis.

Problem

To meet the overall objective of the study an artificial learning problem that could be used as an experimental vehicle was selected. The task selected for analysis and training by experimental training conditions was defined by describing a criterion-performance test. That is, the first step was to describe in measurable terms the performance that we wanted to obtain by training. The goal of the experiment then became to foster the realization of this criterion performance in the least possible time. Learning time rather than error measurement was chosen as the dependent measure, simply because time is a more appropriate measure for a training research study; error measures would have been appropriate for a learning study.

The task defined by the criterion-performance test was one for which the initial repertoire of the subjects used in the study was known. This made it possible to define the training problem as a difference between the relevant initial repertoire and the criterion performance. The method of the study was to analyze this difference to identify important component learning problems, and then to construct experimental training conditions that could be expected to foster fast learning if the analysis were a good one. Such experimental training conditions were developed and used to train subjects until they could demonstrate satisfactory performance on the test that defined the training task. The time required to achieve this training was measured, and the utility of the experimental training conditions was demonstrated by comparing the training time under experimental training conditions with the training time required when a conventional approach to training was employed for the same task.

This kind of study could, of course, be set up for any number of different training tasks defined in terms of a difference between an intial repertoire and a final performance test. By selecting different tasks defined in this way one could expect to discover different ways to analyze the tasks, and different sets of effective training conditions. As will be seen in the analysis presented below, the analysis of a training problem depends upon the availability of information developed in learning studies. It is expected that if many training tasks were studied in this way, it would be found that the various analyses would uncover similar learning problems among some tasks, and that in this way a "catalogue" of learning problems and useful training techniques for each type of learning problem could be developed. In this study, however, only one task was selected for analysis.

Method

The Training Task

The training task chosen for this study required learning the names of the eight electrical circuits shown in Figure 1. The performance test defining the task required the trainee to demonstrate capability to select the correct name for each circuit diagram from a list of all eight names when the diagrams were presented one at a time. More specifically, in the performance test each learner was required to make one complete trial of 16 responses without error—that is, two correct-name responses per circuit. The performance test was administered by means of a Subject-Matter Trainer (see Chapters 23 and 25). This device presented the circuits at an exposure window one at a time. When a circuit was exposed, the learner was required to push a button beside the correct name of the circuit. The names of the circuits were presented on a panel beside the exposure window and

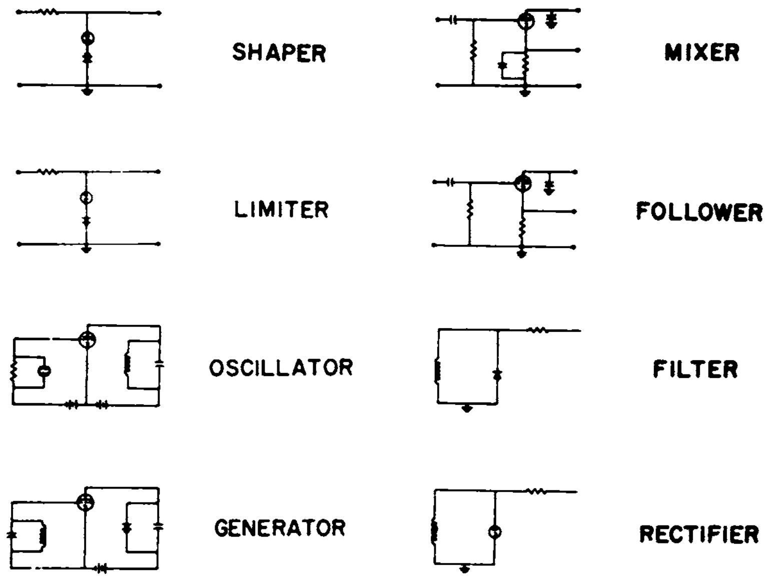


Figure 1. The paired-associate materials used in this study

there was a response button beside each name. If the learner pushed a wrong button, a buzzer sounded; if he pushed a right button, a green light indicated that his response was correct and he was then able to call up the next stimulus item by pushing an "advance" button. The advance button activated a wheel inside the machine which contained the stimulus items, and which rotated to present the next circuit diagram. During the performance test each learner paced his own responses.

Since it was important for this study that learning be defined as a difference between an initial repertoire and a desired performance capability, the subjects were selected so that all had the same relevant initial repertoire. None of the subjects used in the study knew the names of simple electrical symbols, nor were they familiar with the names of any circuit types.

Description of Learning Difficulty

The task defined above in terms of pre-training and post-training performance was not selected for this study until it was determined how difficult it was to learn the task. Measurement of the difficulty was performed in two ways. First, nine subjects were given simple paired-associate training by means of the Subject-Matter Trainer. In this training the circuit diagrams were presented one at a time, and the learner tried to select the correct name for each by pushing the appropriate button on the response panel. From trial to trial the order of presentation of the circuit diagrams was changed, and the arrangement of the eight names on the response panel was also changed. During an initial demonstration trial the subjects were told the correct responses; thereafter an immediate-correction procedure was used.

It was found that it was very difficult to master the task under these training conditions. None of the subjects ever achieved perfect performance. Training was stopped for every subject because of exhaustion. Thus the time spent in training individual subjects varied, depending upon the time of the onset of exhaustion.

The performances achieved on the last two trials by learners trained by this paired-associate method are indicated in Figure 2.

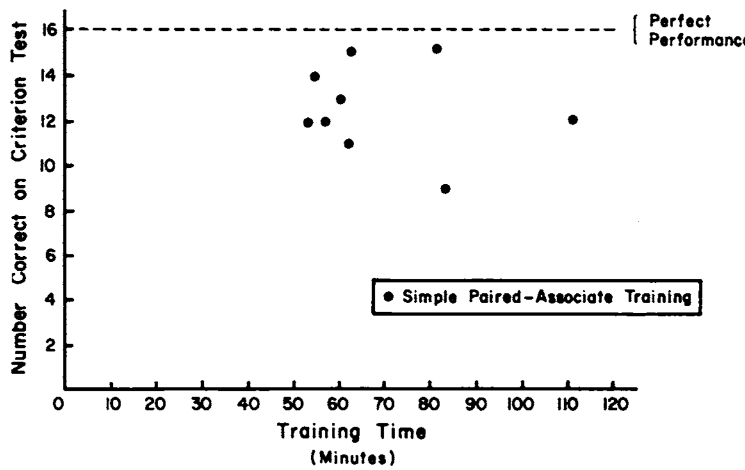


Figure 2. Distribution of training time and test scores with simple paired-associate training

A second experiment to measure task difficulty showed that the task could be mastered under different conditions in a much shorter time. In order to obtain the second measure of difficulty, subjects were allowed to select their own method of study; no formal attempt was made to provide training conditions. Thirteen subjects were requested to study the paired-associate materials for two periods of 20 minutes each. Each subject was given a study card that looked like Figure 1. Each subject studied individually, and each was allowed to use pencil and paper during his study sessions. None of the subjects was given instructions as to how he should study. Each subject was tested after 20 minutes of study, and again after 40 minutes. The result of these tests is presented in Figure 3. The square plots on the left in the figure show the performance of the learners who used self-selected study methods; the plots on the right show the much poorer performance of the men trained by the paired-associate method (as previously shown in Figure 2). The square plots and accompanying frequencies (small numbers beside the squares) at 20 minutes of training time show that three out of the 13 learners achieved perfect performance after 20 minutes of individual study. The square plots (with accompanying frequencies) at 40 minutes show that 11 out of the same 13 subjects had achieved perfect test scores after a total of 40 minutes of individual study.

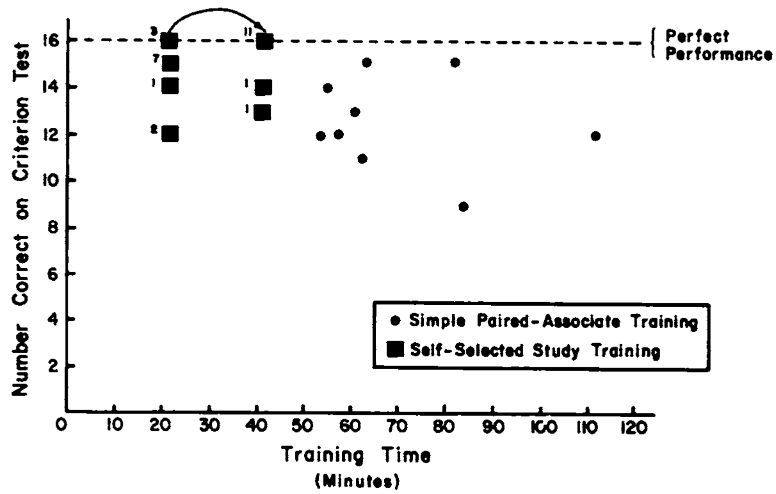


Figure 3. Distribution of training time and test scores for self-selected study training and simple paired-associate training

Analysis of the Task

Given all the above information about the difficulty of the task, the objective of the study became that of devising experimental training conditions that would maximize the efficiency of training in terms of training time. In order to provide a basis for developing these training conditions it was necessary to analyze the training task for the purpose of discovering what learning problems it actually involved.

To analyze the task it was necessary to determine what kinds of learning would be demanded of the learner during training. Thus, the analysis was focused upon discovering what specific learning processes the learner would have to exercise in going from his initial repertoire to capability to perform the criterion task. The answer to this kind of question must, of course, derive from our stock of knowledge about learning processes. Unfortunately, a deplorably small proportion of learning research has been devoted specifically to the development of information about types of learning processes, and any analysis that can be made now is difficult to justify with good research evidence. This kind of analysis requires definitions of the types of learning processes to be identified for developing training materials. Such definitions must, of course, identify each learning process as a process of change.

One kind of learning process that is relatively easy to identify as a process of change is associative learning. When it is defined as a process of change, it can readily be seen that associative learning is involved in the training task that was analyzed. A schematic representation of associative learning defined in this way is presented in Figure 4.

TIME 1 $S_1 \not\rightarrow R_1$ but $S_0 \rightarrow R_1$
and S_1 is not confusable with other stimuli in the task

TIME 2 $S_1 \rightarrow R_1$

Figure 4. Schematic description of simple paired-associate learning as a process of change

The figure is a summary description of the following. At Time 1, prior to introduction to the task, the learner cannot give response R_1 when S_1 is presented. However, at Time 1 the response R_1 is itself in the repertoire, as shown by the fact that it can be given to some stimulus different from S_1. Further, the initial repertoire includes capability to discriminate S_1 from other stimuli in the task to be learned. At Time 2, which is the end of the associative learning process, the criterion performance —capability to give R_1 when S_1 is presented—is within the repertoire.

This definition of simple associative learning deliberately excludes processes which might be called "response learning" and "stimulus learning," because tasks involving these processes require different training conditions from those required in simple associative tasks in which these processes are not involved.

According to the above definition associative learning is involved in the experimental task with which we were concerned in this study. However, it can also be seen, if one attempts to describe the circuit-naming task as a simple associative one, that not all the conditions are met. Thus, although capability to respond to the printed names was present in the initial repertoire of the learner, thus requiring no response learning, the stimulus conditions in the definition can be met only by providing for some kind of stimulus-learning process in conjunction with simple associative learning. Thus, the stimulus terms in this task are clearly confusable with each other at the outset of learning, and this confusability must be resolved before the task becomes a simple associative one. The analysis of the task is not complete with the identification of associative learning as a process required to learn the task; another process must also be involved.

What is the nature of the stimulus problem? From inspection of the stimulus materials it appears that the list is made up of four pairs of terms that may be readily confused within pairs. Empirical observation confirms this. As long as these terms are confused, the learner will not be able reliably to give a unique name for each term. Therefore, before associative learning can take place some sort of stimulus-learning process must take place that will result in the formulation of a unique response of some sort as each term is presented. More specifically, it was reasoned that a unique implicit response to each of the stimulus terms would have to be learned.

The learned implicit response could then function as the non-confusable stimulus event to which each of the names could be associated by a simple paired-associate-learning process (Wulff and Stolurow, 1957). The kind of change process that was thought to be required is described schematically in Figure 5. This stimulus-learning process was identified as the second critical learning process to be fostered by the selection of appropriate training conditions.

TIME 1 S_1, S_2 } confusable   when $S_1 \rightarrow R_1$ and $S_2 \rightarrow R_2$ are both required in the criterion task

TIME 2 S_1, S_2 } not confusable

Figure 5. Schematic description of stimulus learning as a change process

In Figure 5, S_1 and S_2 stand for stimulus terms in the group of stimuli involved in the task. Non-confusability of S_1 and S_2, as called for at Time 2 in the definition, might be demonstrated by capability of the learner correctly to call sequentially presented pairs of the stimuli "same" or "different."

No claim is made that the above analysis is the best one possible; it is simply the best that was devised. At the time when the analysis was made, it was expected, nevertheless, that it would be a large enough step in the right direction to make it possible to design a good training procedure that would exemplify the usefulness of the analytic procedure.

Training Materials and Administration

The training materials that were developed for experimental training were, of course, based upon the analysis described above. In preparing the training materials, certain restrictions were observed. It was agreed that whatever training conditions were set up would be wholly determined by the analysis (that is, it was agreed not to use any mnemonic devices, for example), and it was decided that whatever training procedure was established would be one that could be standardized (that is, it would be suitable for machine presentation, if desired). Mnemonic devices were ruled out because it was felt that no new information would be gained through the use of such procedures.

In order to develop training conditions that would facilitate the learning processes identified in the analysis, it was necessary to consider the order in which the two learning processes should be fostered during the experimental training. Since paired-associate learning is contingent upon the employment of a reliable stimulus event by the learner, it was obvious that stimulus learning should be promoted first. Therefore, training conditions were selected for the beginning of experimental training that would foster the required stimulus learning by utilizing performance capabilities already in the repertoires of the learners. A card-sorting procedure that used the materials shown in Figure 6 was selected for this purpose.

Training materials consisted of a set of 48 cards with a drawing of one of the eight circuit diagrams on each card, and of the sorting boxes shown in the figure. Each sorting box was labeled with one of the circuit drawings. Trainees were required to sort the cards into the boxes. In the first phase of sorting (as shown in the figure by Boxes 3, 4, 7, and 8), flaps that could be used to cover the circuit-diagram labels were raised on all boxes and the cards were sorted into the boxes by matching each card with the correct label. At first this could only be done slowly, though it was done reliably. Ability to sort the cards rapidly into the correct boxes was developed with practice, and was taken as evidence that the stimulus terms were no longer confusable.

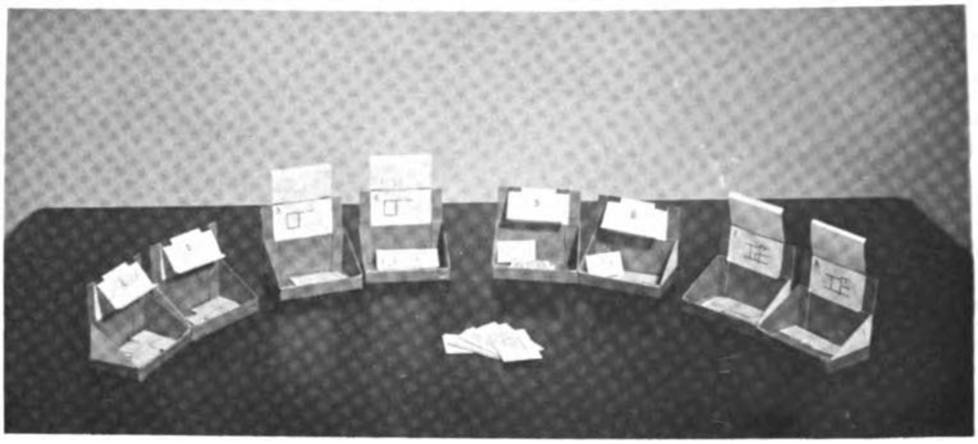


Figure 6. Materials used in sorting-training arranged to show the procedure for both the first and second phases of sorting

At the conclusion of the sorting-training the confusable pairs of circuits were no longer confusable, and the conditions for initiating simple paired-associate training were thus established. Paired-associate training was introduced by taking advantage of a by-product of sorting-training. That is, during sorting-training the subjects also learned to associate the position of each sorting box with its circuit picture. To capitalize on this, paired-associate training was introduced by continuing the sorting procedure with the labels (circuit diagrams) on the boxes covered by flaps numbered to correspond to the box positions. This continuation of the sorting procedure required each subject to learn to associate a number response with each of the circuits. At the completion of this phase, associative training was then continued under quite different training conditions.

During the next phase of associative training all of the correct responses (circuit names) were displayed on a panel, and the learner was required to identify the correct name as each circuit was shown through an aperture in the panel. As each circuit was shown for the first time, the correct name was given by the experimenter. On subsequent trials, the subject was required to give the correct number for each circuit and to select the correct name. A gradual, cumulative training technique was employed. That is, training was initiated using only two paired associates. Then, when capability to select the correct name for the two had been demonstrated, the third paired associate was added, and so on until all eight pairs were included. Only after the learner had demonstrated capability to select correct names for all eight pairs with this technique was he tested using the standard performance test that defined the objective of the experimental training. The total time required to accomplish the sorting and the paired-associate training was recorded for each subject.

The critical feature of this "engineered" paired-associate training technique was the use of the easily learned number response to support the learning of the circuit name. Thus, the association of the number response with the circuit drawing and with the name was promoted to provide an alternative route to the correct name, which could be used by the learner if he were unable to recall the name itself. This supportive association was introduced to increase the probability that correct-name responses would be made to the relevant stimuli early in learning, thus providing an opportunity for positive reinforcement to act to strengthen the correctly practiced associations.

Subjects

The subjects were Air Force technicians selected from a technical training school. They were trained independently, using the materials and procedures described above.

Results

The performance of learners trained with the experimental ("engineered") training conditions is summarized in Figure 7. It can be seen that all of the learners demonstrated the required performance capability when they were tested on the Subject-Matter Trainer. It can also be seen that, although the time required for mastery ranged from 12 minutes to 27 minutes, the training time is considerably less than that needed by the trainees who did all of their learning on the Subject-Matter Trainer. Comparison with the self-selected study-method group is not especially useful since the self-determined training conditions used by each trainee are unknown, and probably differ from learner to learner anyway.

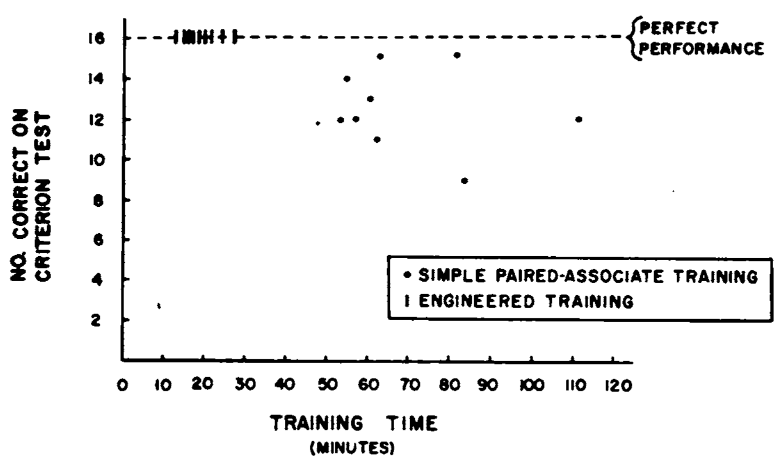


Figure 7. Distribution of training time and test scores for experimental training and simple paired-associate training

It was found to be quite difficult to devise training conditions that would readily support the specific kind of learning activity prescribed on the basis of the learning analysis. As a matter of fact, several alternative methods were tried, but none other than the one described above proved suitable for the majority of the trainee population. Throughout the try-out period, however, the training conditions that were tried were based on the same analysis as that reported here. All the training conditions tried were successful to some extent. The principal difficulty was in finding training conditions that would work for all subjects. Most of those tried worked well with some learners and poorly with others.

It is reasonable to question whether or not the effectiveness of the training procedure that was used can be attributed to the analysis of the task and to the training conditions employed. In order to obtain checks on the analysis, two auxiliary studies were performed.

As a check on the guess that one of the critical processes was simple discrimination learning independent of name learning, trainee responses to the pictured circuits _per se_ were compared before and after nomenclature learning. Subjects in this experiment learned the names of half of the eight circuits (two pairs of confusable diagrams) by a simple paired-associate technique. Following this training, each of the eight circuits was tachistoscopically displayed one at a time at near-threshhold levels, and subjects were required to identify each displayed circuit by pointing to the same circuit on a card containing all eight circuits. Mean recognition performance for the unfamiliar circuits was one item better than chance; mean recognition performance for the circuits for which names had been learned was over six items better than chance. The difference between performance on the familiar and unfamiliar circuits was reliable at the three per cent level of confidence. If stimulus learning were no problem in this task, no difference in recognition performance would be expected. The implication of this finding is that learning relevant to the stimulus object _per se_ takes place during nomenclature training.

A second auxiliary experiment was performed to check the guess about the requirement for associative support during practice on this task. In this experiment learners practiced using a paired-associate technique that was, in all respects but one, like that described previously for the subjects given all of their training on the Subject-Matter Trainer for the purpose of determining task difficulty. During the initial stage of this auxiliary experiment the eight stimulus terms were presented in a fixed order, and the response terms were maintained in fixed positions. Maintaining fixed order and fixed positions were expected to provide associative support through position cues. Subjects practiced under this condition of training until they had over-learned the responses. In the second stage of training, these subjects practiced under the same random order of presentation conditions used for the first reported group that learned on the Subject-Matter Trainer. The performances of these two groups trained on the Subject-Matter Trainer are compared in Figure 8. As expected, the subjects trained under conditions that initially afforded associative support learned the task more quickly. This finding supports the idea that redundant associative connections can be used to support associative learning.

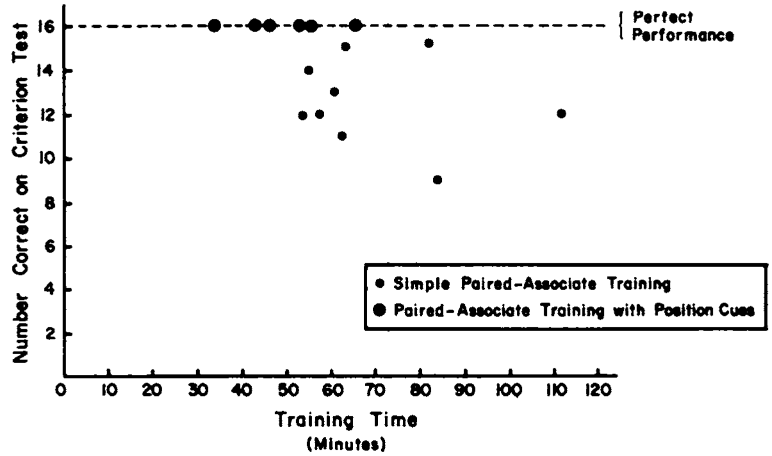


Figure 8. Distribution of training time and test scores for paired-associate training with position cues, and simple paired-associate training

Summary

This study exemplifies the use of one strategy for deciding upon the characteristics of an effective training program, and it illustrates the utility of that strategy. Briefly, the strategy is based on recognition of the fact that a training situation must have two characteristics: (1) it must provide opportunity for the learner to make correct responses to relevant cues; and (2) it must provide training conditions that will foster the processes of change that must take place during learning. To apply the strategy, opportunity for practice of correct responses to relevant cues must be provided as defined by the criterion performance. Over and above this, one must identify the initial performance capability of trainees, and determine by analysis what learning processes are required to carry a learner from his initial capability to the final performance required. Finally, training conditions that will foster each of the identified learning processes must be selected and arranged in an appropriate order to provide for maximum training efficiency. The analysis of the learning situation presented above exemplifies a way in which "what is learned" can be identified by analysis of the training problem. The experimental training program described demonstrates that "how it is taught" can be derived from consideration of "what is learned." The results of the main study demonstrate the efficiency with which training can be conducted when this relationship is taken into proper account. Finally, additional confidence in the learning analysis is afforded by the results of two supporting experiments. The implication of this study is that "what is learned" can usefully be described in terms of the learning processes that must be employed during mastery of the task, and that "how it is taught" can best be determined by consideration of what kinds of training conditions will facilitate the critical learning processes identified by such an analysis.

CHAPTER 31

SOME CONCLUSIONS CONCERNING STUDENT RESPONSE AND A SCIENCE OF INSTRUCTION

A. A. Lumsdaine

The basic conception that lies behind the preparation of this volume and the experimental studies on which it is based is that the development of a science of instruction—or more particularly a science of programmed instruction—is desirable and feasible. This basic notion, further discussed later in the present chapter, holds that not only learning but also instruction—the management of learning—can be systematically described and improved through experimental inquiry (cf. also Lumsdaine 1962b, and Lumsdaine and Glaser, 1960, p. 564).

A further assumption is that this experimental inquiry is likely to be most fruitful when it is guided by, and interacts with, the development of appropriate theory. Such theory may include not only attempts at a single integrated theory of instruction, the full development of which would appear to be premature at this time, but also specific sub-theories. Such sub-theories may attempt to develop, for example, testable implications of the role of implicit responses in paired-associate learning (cf. Chapter 22), or, as developed by Sheffield in Chapter 2, theorization may be addressed to problems in the structuring of demonstration and practice for the learning of complex sequential procedures.

As was indicated at the outset in Chapter 1, a further conviction which has given direction to most of the papers included in this volume is that an important characteristic of successful instructional theory (and theory-based experimentation) is that it is conceived with particular reference to the role of student response—implicit or explicit—and to the control of stimulus conditions that govern such response. Of particular interest are explicitly occasioned responses by the student, whether overt or covert, since these can be related more directly to antecedent stimulus variables than can implicit responses for which the occasion is an unanalyzed complex of concurrent and preceding stimulus events. Finally, overt responses by the student, regardless of whether they are used advantageously to improve instructional efficiency (cf. Chapters 18, 27, and 28), are of special interest because of the information they provide to the experimenter or programmer of instruction with respect to relevant aspects of learner behavior.

This chapter will offer some comments on implications of the work here reported for a science of programmed instruction, and will suggest some of its relations to current conceptual and experimental problems in this field. In doing so, no attempt can be made to deal systematically with the rapidly expanding volume of experimental work that has appeared during the last year since publication of Lumsdaine and Glaser's 1960 source book, Teaching Machines and Programmed Learning (here abbreviated TM&PL). The number of papers that have appeared and the rapidity with which new ones are appearing would alone make this unfeasible within current space limitations. Thus, the attempt is merely to relate some of the implications of the earlier work reported herein to some current directions of experimental effort, citing occasional recent experiments of particular relevance.

As noted in Part IV of TM&PL, the historical stream of work of which these Air Force studies form a part developed, to a very considerable extent, quite independently of Skinner's work at Harvard, though perhaps motivated largely by the same *Zeitgeist*. The notion that student-response variables are of central importance arose in this case through experimentation with "programmed" instructional media in audio-visual form. This orientation, as it affected most of the papers reported herein, grew largely out of the initial experiment on student response (or "audience participation") by Hovland, Lumsdaine, and Sheffield (1949, Chapter 9), described in Chapter 1. A considerable number of individuals working with programmed instruction have heretofore been largely unacquainted with the details of this converging stream of work on student-response procedures used with audio-visual media. It is hoped that one of the contributions of the present volume will be to remedy this hiatus, which is readily understandable in view of the fact that the large majority of these papers have heretofore been published only in the form of mimeographed or multilithed reports that have not been generally accessible.

A number of the studies that have been presented here are of interest from both theoretical and practical points of view. Most of the studies were done in the context of research on instructional films. Their relevance to self-instructional media arises not only from the general communality of instructional variables which would be expected to operate in any form of sequenced instructional presentation, but also from the fact that, in nearly all of these studies, conscious attention was paid to the guidance of student responses through appropriate stimulus techniques. In most of the studies, furthermore, provision was made—as is done extensively in current self-instructional materials—for frequent overt student response, with immediate feedback in the form of correction/confirmation or knowledge of results.

In attempting to identify some of the common threads and implications discernible in these quite varied research studies, it is convenient to adopt as a framework a few gross categories which, though imperfectly defined and partly overlapping, can serve to help structure the discussion. The following four categories are used for this purpose:

Cue Factors. These involve means directly calculated to improve the accuracy or appropriateness of student response during the course of instruction. Such factors include: prompting or use of discriminative cues for guiding implicit as well as overt responses; the use of prompts as related to the size of steps in a program; direct variations in "step length," as in the procedural-demonstration studies of Part I; and other factors that relate to step size and error rate in student practice.

Also relevant here are perceptual factors that are related to these objectives even though they are not tied directly to occasions for explicit responses during instruction (e.g., Chapters 10 and 17).

Transfer Factors and "Vanishing." Here the concern is with factors in the acquisition situation which might be regarded as comprising a sub-category of cue factors, but which particularly affect the transfer from conditions of practice to those of performance (i.e., behavior in the application or terminal, unprompted situation). Included here are the problems of "fading" or "vanishing"—i.e., how to assure the continued elicitation of responses as prompts are withdrawn, thus mediating transfer to the test or terminal-performance situation.

Response Factors. These relate to the nature and role of the students' responses and their consequences, including the questions of overt versus covert responding, other aspects of the form of response, and the role of confirmation-feedback-reinforcement. It is to be noted that the latter often provides an occasion for further correct implicit practice, as well as providing pure "reinforcement" in the sense of reward which motivates the student to continue paying attention and keep on working at the instructional task. Closely related is the question of the function of explicit student response (and attendant confirmation/feedback) in fostering generalized or motivational effects, as distinguished from sheer practice effects on the material specifically rehearsed (or on performances closely adjacent on a gradient of stimulus or response generalization).

Organizational and Progression Factors. These factors relate to the organizing and sequencing of instruction. Although such factors also obviously affect the probability of correct response (and hence interact with step size and the use of prompts), it is convenient to distinguish organizational factors as a separate category. Considered here are patterns of repetition, placement of review, and other factors of sequencing that enter into total instructional organization (as distinguished from factors that are more evident at the molecular level of operation—e.g., within a particular instructional frame or segment).

Some comments concerning factors in each of these four categories are presented in the following sections, followed by some concluding remarks on some characteristics of a science of instruction that might deal systematically with such factors.[1]

[1]A further contribution might be made here by examining in detail some of the methodological problems and difficulties which have either been successfully avoided by appropriate design and procedure in the studies reported herein, or which in some instances have lessened the sharpness of the implications which could be drawn from the findings. A detailed discussion of methodology is considered beyond the scope of the present volume, but a more systematic account of methodological problems in research on programmed instruction generally (and in investigation of student response as a factor in such instruction, in particular) is being prepared separately under a contract with the Educational Media Branch of the U. S. Office of Education.

Cue Factors

The function of cue factors, as the term is used here, is to guide the learner to make correct responses during instruction.[2] These responses may be either explicitly directed or of an implicit, perceptual nature. The manipulation of prompts to improve the correctness of overt student response during instruction is perhaps shown most conspicuously in the study by Kimble and Wulff (Chapter 15), and in the studies of prompting as compared with confirmation procedures reported by Briggs in Chapter 23 and by Cook and associates in Chapter 21. Also relevant to prompting, though perhaps less obviously so, is the use of serial-order cues in early trials on paired-associate material reported as an ancillary study by Wulff and Emeson in Chapter 30.

The studies by Maccoby, Sheffield, and associates reported in Chapters 3, 4, and 5 dealt, it will be recalled, with size of the demonstration units which students viewed prior to practice. These studies relate to organizational factors, discussed later, but also can be viewed as experimental variations of the directness of prompting or cueing provided to the student, since, in a procedural task, the use of very short demonstration segments provides cues that are temporally less remote (and also less subject to interference from intervening material) than in attempted practice after lengthier segments of demonstration. Attention to the use of special prompting devices with particular reference to the guiding of implicit responses is the concern of the studies reported by Wulff and Kraeling in Chapter 10 and by Lumsdaine, Sulzer, and Kopstein in Chapter 17.

Optimal Use of Cueing

The study by Kimble and Wulff (Chapter 15) affords a rather clear experimental demonstration of the superior effectiveness of additional prompting as compared with less prompting. Current discussions of auto-instructional programming (e.g., Klaus, 1961) stress not only the value of adequate prompting but also the importance of not over-prompting the student. It is of interest that the cueing or "response-guidance" techniques reported by Kimble and Wulff in Chapter 15 sought to foster current responding without complete prompting that would have "given away" the answer completely and reduced the learner's task to mere copying. That the latter form of "over-prompting" would have been undesirable was simply assumed by these investigators; it would have been of interest—particularly in the light of the later data by Cook and others favoring straight prompting—if a comparison had been made between more complete prompting and the partial prompting afforded by Kimble and Wulff's response-guidance techniques. Even more interesting would be a treatment in which the directness of prompting is progressively decreased, as suggested later in relation to the discussion of vanishing.

[2] In this discussion the terms "cueing" and "prompting" are used more or less interchangeably, without reference to the terminological distinction between the two proposed by Briggs (1960).

The use of prompting techniques for mediating purely verbal paired-associates material appears to have had relatively little attention in the general literature of experimental psychology. An exception is an early experiment on the use of semantic prompts in verbal learning by Pan (1926), who reported that semantic hints indirectly suggesting the response term facilitated the learning of arbitrarily paired words. The question of optimal degree of prompting in learning paired associates (or other material where each item has to be repeated several times) involves not only the degree of prompting used on any one occasion but also the frequency with which prompting is employed on repeated exposures. The limited evidence on this point comes primarily from the comparing of the results of Cook and collaborators (Chapter 21) with those of Angell and Lumsdaine (Chapter 24), and is discussed below in connection with the questions of vanishing and transfer from instructional to test situations.

The Cook results, taken by themselves at least, militate for stress on maximum prompting as a way to ensure correct responding and thereby to mediate more effective acquisition. The same is true of the results presented by Briggs in Chapter 23. The use of overlapping versus simultaneous presentation of stimulus term and response term in paired-associates learning, which entered as one of the variables under study in the experiments of Kopstein and Roshal (Chapter 20), also can be considered to support response prompting, since the simultaneous pairing presented a full, immediate prompt (as contrasted with the less effective staggered presentation, in which the stimulus term was first presented unaccompanied by the prompt of the response term).

Cueing as Error Control

In discussing their results, a significant emphasis concerning the use of prompting is pointed up by Kimble and Wulff (Chapter 15) by their finding that the effects of prompting or "response guidance" were accounted for almost entirely in terms of the elimination of common errors which their particular prompting ("response-guidance") devices were designed to prevent. They draw the conclusion that the effective use of student-response procedures derives from the fact that incipient mistakes are corrected before they are made. This emphasis on error-control in turn implies a requirement for identifying those specific errors which students are most likely to make. At present such error tendencies generally need to be identified empirically, on the basis of student response to preliminary forms of an instructional program; however, a well-developed science of instruction would eventually include bases for predicting specific types of errors on theoretical grounds.

It may be noted that current emphasis on low error rates in self-instructional programs, achieved by this kind of tryout and revision, has sometimes assumed low error rates to be desirable without experimental documentation that the measures taken to reduce error rate during practice (instruction) will also benefit terminal performance. The Kimble and Wulff study, however, showed directly that prompting devices to control errors during practice did in fact produce better performance in the test situation following instruction. The same was true of the studies of Maccoby, Sheffield, and collaborators, though it will be recalled (cf. Chapter 5) that the most highly cued (shortest demonstration-segment) treatment was accompanied in some cases by a decrement in performance in the unprompted test situation.

Cueing and "size-of-step"

The question of prompting is intimately related to the notion of using small steps in an instructional program. Variant meanings of the latter notion have been indicated previously by Lumsdaine (1959b). For example, distinctions need to be made between the "step-size" characteristics of length of frames, length or complexity of the required responses, difficulty of the response (as judged a priori or as measured objectively), degree of molecularity into which the elements of a subject matter are broken down, and number of responses (or complexity of response) before correction/feedback or reinforcement is introduced. In the case of procedural learning, where a demonstration is used as a basis for imitative practice, as in the studies reported in Chapters 3 through 10, the application of the notion of size-of-step applies quite literally, referring to the length of the demonstrational segment that the student watches before attempted practice.

It may be observed that the relationship between prompting and the several aspects of "size-of-step" in instructional programs is one in which cart and horse may sometimes seem to be reversed. In general, one would say that steps are made easy (or, in this sense, "short" or "small") when ample prompting fosters correctness of response. Conversely, however, the physical length of a self-instructional frame or sequence may contribute in effect to prompting through reduction of interference from other material which would represent a distracting element in a longer instructional segment. This is particularly evident in the case of procedural learning by imitation (cf. Part I), where the use of a short demonstrational step provides a direct, easy prompt for the imitation of that step. In the two studies summarized by Maccoby and Sheffield in Chapter 5, the use of such "small steps" led to generally superior performance in practice and in some instances to superior later performance under unprompted test conditions. In addition, attention to perceptual-integration factors (cf. Chapter 2) and the use of context cues in the designing of a demonstration or other instructional sequence (cf. Chapter 9) can afford one basis for effective prompting.

Cueing of Implicit Responses

With the possible exception of instances in which very short, discrete (unitary) responses are practiced explicitly an item at a time—as in the simplest forms of paired-associates or serial learning—it is evident that much of the learning from academic instruction is mediated by implicit responses generally not identified in detail. It is also clear that we have at present only a very limited understanding of the nature and operation of the implicit responses involved in reading, listening, or watching a demonstration as a basis for later practice or performance. Nevertheless, we are evidently entertaining some provisional hypotheses about the functioning of implicit responding when—as in the studies reported in Chapters 10, 11, and 17, for example—we manipulate cue features of an instructional presentation that are not directly followed by overt, explicit responses but rather are expected to carry over into later performance by some form of verbal or perceptual mediation.

For example, the effect of familiarization training as used by Wulff and Kraeling (Chapter 10) presumably involves perceptual discriminative responses which might sometimes take the form of verbally and perceptually mediated distinctions—e.g., "the washer with the small hole in the large disc goes next to the clamp," etc. Similarly, the use of attention-directing arrows and labels studied by Lumsdaine, Sulzer, and Kopstein (Chapter 17) provides additional cues above those provided by the verbal material in the narration—e.g., for such a response as the student saying to himself: "These two little lines mark off two spaces which each equal 25 thousandths, and so, corresponding to them, I should write .050, thus... to which I must then add in the space below the further reading given up there by the barrel-scale setting."

Lack of an adequate understanding of the operation of mediating responses severely limits the formulation of hypotheses concerning the properties that make for effective prompting of such responses, and forces the designer of instructional materials to proceed on a largely intuitive basis. One of the problems which arises in doing so is the potential conflict between competing requirements of the total instructional situation. Such problems often can be conceived of in terms either of response-cueing or of the organization of instructional sequences. For example, the procedure of introducing familiarization with components just before each component was to be added to an assembly—found by Wulff and Kraeling (Chapter 10) to be superior—offers the advantage that appropriate discriminative responses are set up in temporally close proximity to the further responses of part selection and manipulation which they are helpful in mediating. On the other hand, interruption of the assembly task to point out these discriminating features constitutes a digression in the smooth progression of the assembly act and may interfere with its being learned as an integrated unit. It is of interest in this connection that the advantage of the "*in situ*" familiarization procedure in Chapter 10 was shown at a significant level only for reduction in *selection* errors; lack of a significant gain with respect to assembly errors may have resulted in part from interference with perceptual responses conducive to task integration. On the other hand, for the kind of prompting cues afforded by the animation devices used in teaching the essentially perceptual discriminations involved in micrometer reading (Chapter 17), this disadvantage presumably was not incurred, since here the prompting cues were used to foster implicit responses which were indigenous to the principal response pattern to be learned.

The mediation of responses utilized in the form of delayed imitation required in serial-construction or assembly tasks (e.g., Chapters 1 through 11) involves the interesting feature that, in addition to purely symbolic or perceptual responses, we may presume the occurrence of incipient motor responses anticipatory of the later-required overt performance. Roshal's general hypothesis (Chapter 11) was that instruction would be most effective as the demonstration approached a representation of the learner himself performing the task. This hypothesis (cf. also May, 1946, 1958) perhaps points most directly to a transfer requirement for similarity between the cues used in training and those to be present in the application or testing situation, as further discussed in the next section of this chapter. However, the argument for use of motion and the "subjective" camera angle could also be stated in terms of their providing effective prompts for appropriate implicit responding while watching the demonstration.

It should be recognized that it is often very difficult to identify the cues which actually serve as prompts for implicit as well as overt responses in a program for teaching a complex subject matter (and even in the simpler paradigm of paired associates). The detailed analysis of cues in programmed material by Meyer (1960) is instructive in this respect. It may also be noted that, where explicit response-prompting cues are used, their effectiveness (and their importance—in the sense of how much need there is for them) may be dependent on the effectiveness of other, less explicit or less formalized prompting procedures that are used. For example, the confirmation or correction sub-frame in a self-instructional program, and the exposure of the response term in a paired-associate list, in themselves provide prompts for correct responding at least at the implicit level—even when there has been previous and immediately preceding opportunity for overt response. To the extent, however, that adequate prompting was provided for an overt response, the efficacy of (and need for) the further implicit-response prompting provided for by the confirmation panel would be expected to diminish, as indicated in data recently reported by Angell and Lumsdaine (1961a).

As an example of the interaction among classes of instructional variables it may be noted that the prompting issue is intimately related to current controversy and to past experimentation—including here the Michael-Maccoby study (Chapter 18), the McGuire study (Chapter 26), and the Kanner and Sulzer study (Chapter 28)—on overt versus covert responding. The covert-response conditions employed in these studies prevent the learner from making an incorrect overt response and may serve to maximize the proportion of responding that is under the control of a prompting cue—that is, the implicit responding which can occur after "KCR" or "confirmation" is provided.

Response Guidance through Prior "Set-Producing" Instruction

A special aspect of the guidance of student response that is not so readily conceived within the framework of direct prompting is the use of prior instruction and attention-directing devices to structure subsequent implicit (or overt) responding by the student. Examples are found in the use of pre-instructional tests or introductory "orientation" that stress what is important to attend to in later instruction. A special case is the form of instruction utilized by McGuire (Chapters 12-14) in his film demonstrations preceding overt practice. Direct coaching right at the time of practice can undoubtedly be more effective when feasible; e.g., in the Wulff and Kraeling study (Chapter 10), familiarization with critical cues just preceding each step of practice was more effective than similar familiarization given earlier. However, prior instruction and guidance preceding overt practice by an appreciable interval might appear necessary where the nature of the task makes it difficult to provide guidance in situ—e.g., in motor skills not amenable to direct, step-by-step imitation.

Another situation in which prior guidance rather than point-by-point prompting may be of value is one in which a general "set" mediated by prior instructions can furnish direction in common to implicit responses that will be appropriate throughout a considerable block of succeeding instruction. The effectiveness of tests preceding an instructional unit for this purpose is investigated by several studies summarized in the Appendix—e.g., Miller and Klier (1956a), Kimble (1955b)—as well as being evidenced in the contribution of the pre-instruction test used by Lumsdaine, Sulzer, and Kopstein (Chapter 17).

A fairly common finding is that prior attention-directing devices may have both positive and negative effects. For example, the use of slow motion in McGuire's study (Chapter 13) made a positive contribution to the learning of the selected postural items for which it was employed, but at the expense of reduced learning of other points for which slow motion was not used in the film. Similarly, in McGuire's experiment the contributions of the narration to items to which it directed attention was also effected at the expense of reduced learning of other material not similarly stressed. These findings parallel results for substantive verbal learning earlier reported by Romney (see Appendix) and by May and Lumsdaine (1958, Chapter 7), who found that attention-directing pre-instruction tests produced gains in the aspects of the content thus stressed, but with an accompanying decrement in the material to which attention was not directed by the pre-film instruction.

The "set-producing" effect of pre-instruction tests can be differentiated from the use of the prior announcement of a post-instruction test, as employed by Hovland, Lumsdaine, and Sheffield (1949), in the second study reported in Chapter 19 by Maccoby, Michael, and Levine, and (unsuccessfully) in Michael and Maccoby's study in Chapter 18, to manipulate level of motivation or attentiveness. Non-uniformities in the effectiveness of such procedures, rather than being the occasion for despair about the generality of findings, would appear to suggest the need for contingent generalizations on which the effects of such procedures are predicted differentially in relation to task and situational variables. The problem then is to identify the circumstances under which the procedure will and will not be expected to be effective. As is evident in the situation encountered by Michael and Maccoby, such predictions may be useful not only for the engineering of instruction, but also in the design of experiments where it is desired to introduce differential levels of motivation for heuristic purposes.

Transfer Factors and Vanishing

The central concern of this section is with stimulus factors during instruction which foster correct responses in the later test, application, or transfer condition—i.e., with factors which foster transfer from acquisition to conditions of desired terminal behavior. A sub-category of special interest here concerns the question of "vanishing" or "fading" of the prompts which are used to promote correct responding during initial stages of learning. Other factors of interest include certain perceptual elements that affect response both in the training situation and in the later testing situation.

Evidence on the Effectiveness of "Vanishing" and Partial Cueing

Despite the relative paucity of studies employing explicit experimental manipulation of prompting factors within the context of self-instructional programs, there is more experimental evidence on the usefulness of prompting as such than on the value of "vanishing" or weaning the student from dependence on prompts. This applies both in the studies reported in this volume and, thus far, in subsequent experimentation on self-instructional programs.

Evidence bearing on the rationale for vanishing is of three kinds. The first and more direct compares conditions in which full prompting is used throughout with conditions in which initial strong prompting is later diminished in frequency or strength. The main example of such a study in the present literature is that reported by Sheffield, Maccoby, and collaborators in Chapters 3 through 5. (See also summaries of further exploratory experiments by Angell and Lumsdaine (1961 a,c) and by Guthrie and Lumsdaine (1961) presented in the Appendix.) In a second kind of comparison, some form of vanishing is used but the comparison condition is something other than continued use of full prompting. Examples are found in the present volume in the second study reported by Briggs in Chapter 23 and in the ancillary study by Wulff and Emeson (Chapter 30) in which initial prompts provided by a fixed serial position for paired-associate items were later withdrawn. In the latter study, though vanishing was employed, the nature of the comparison was really such as to demonstrate the value of the early prompting rather than of the later vanishing of prompts. A third kind of study is typified by the 1960 Angell and Lumsdaine experiment (Chapter 24) and other studies of partial prompting. In such studies, although there is not progressive diminution of prompting as the instructional sequence progresses, implications of a general partial-cueing rationale underlying vanishing are nevertheless tested.

It is worth reviewing the evidence on the advantage of vanishing in procedural learning obtained by comparing a transition from more prompting early in training to less in later stages (versus continued use of full prompting) in experiments summarized in Chapter 5 by Maccoby and Sheffield. It will be recalled that in their "transition" condition, the learner practiced after each short segment of demonstration on the first trial, then after longer segments on the next trial, and finally after seeing the entire demonstration as a whole. For maximizing transfer from training conditions to conditions of criterion performance without demonstrational support, the data indicated that this transition method was superior not only to methods that provided less direct prompting initially but also to continued use of maximal short-step prompting. For this kind of demonstration-imitation learning the directness of prompting can, as noted previously, be considered inversely proportional to the length of the demonstration steps presented before the student attempts to practice what he has seen. Thus the application of the concept of vanishing is a fairly straightforward one when one follows the procedure of progressive increase in the length of demonstrational steps at successive stages of practice—regardless of whether this is automatically regulated by the experimenter or program, or whether the learner himself adjusts the length of demonstration in an attempt to perform as long a segment of the operation as he is capable of at any particular stage in practice.

The second of the studies reported by Briggs in Chapter 23 compared several transition conditions that in a sense amounted to various conditions of vanishing, rather than comparing a particular transition or vanishing condition against continued use of full prompting. (The latter condition, used in the first of the two studies described in Chapter 23, could not be directly compared with the transition conditions because the two experiments were conducted at different times with different subject populations.) The three transition conditions compared amounted in effect to: (1) full prompting for 10 minutes followed by an "intermediate" level of prompting for 10 minutes; (2) full prompting for 10 minutes followed by minimal prompting for 10 minutes; and (3) full prompting for five minutes, then intermediate prompting for five minutes followed by 10 minutes with minimal prompting. Differences obtained between these three conditions were non-significant. However, had significant differences among these three conditions been obtained they would merely have shown that one "gradient" of prompt withdrawal was more efficient than a different gradient, for this particular task, unless there were a theoretical basis for predicting superiority of some particular gradient in more general terms. The fact that the third condition used more stages of gradation would not necessarily predict it to be optimal since, for example, longer-continued use of maximal prompting might have been called for.

Although the response guidance provided by Kimble and Wulff (Chapter 15) was a form of partial rather than complete prompting, unfortunately no comparison with variation in degree of prompting was made. As noted previously, it would be interesting to repeat an experiment similar to that of Kimble and Wulff in which the directness of prompts was progressively decreased. These authors point out that, with a more complete guidance technique instead of the one that they used—that is, an arrow pointing directly to the correct response, rather than one of the forms of bracketing or approximate response localization which they employed—the learner would always make correct responses, but the cues that he would use would be different from those he would have to employ in a test situation. However, Cook's results (Chapter 21) would suggest that this might be less disadvantageous than supposed, and even the vanishing concept would predict that such very direct prompting should be used initially, followed by less and less direct forms of response guidance.

In their comparison of training procedures termed "prompting" and "confirmation," the "confirmation" procedure reported by Cook (Chapter 21) provides either verification or correction of the response that the student has given. With both procedures, an unprompted and unconfirmed test trial was given by the experimenters after every three training trials. The condition using prompting trials was clearly superior to the confirmation condition, both in the original experiment and in later varied replications reported by Cook and his associates, as well as in the experiment reported by Briggs (Chapter 23).

The general argument for the principle of "vanishing" suggests that, regardless of specific temporal gradient, some <u>mixture</u> of prompted trials (to get the correct response elicited) and unprompted trials (to give practice in the unprompted test situation) would be a better condition for learning than all unprompted <u>or</u> all prompted trials. As pointed out by Angell and Lumsdaine (Chapter 24), this mixed condition was in fact actually present in the prompting condition used in experiments by Cook and others (Chapter 21) because of the use of frequent test trials. A mixture of prompted and unprompted trials could, integrated over time, be considered as one form of partial prompting (as compared with complete prompting). The basic comparison in the Angell and Lumsdaine experiment reported in Chapter 24 was between two groups, one of which received the prompting on all trials in learning a set of paired-associates, and one of which practiced responding without prompting (but with confirmation) on every fourth trial. It will be recalled that the results, for learning materials and time intervals substantially the same as those employed by Cook, indicated that learning was more efficient under the condition of incomplete prompting (prompting on three-fourths of the trials) than under conditions of complete prompting (prompting on every trial). This conclusion is in line with the general theoretical argument supporting vanishing; even though progressive diminution in the amount of prompting was not given, the principle was involved in that some practice was given, during the instructional sequence, in responding without prompting.

A more specific theory of vanishing which is suggested by Angell and Lumsdaine has been schematized in graphical form by Lumsdaine (1959b). This theory is an extension of the notion of vanishing as propounded by Skinner and other writers. It not only asserts that prompts should be withdrawn, but further asserts that, at any point during acquisition, the strength of unconditional cue or prompt provided to the learner should satisfy the dual conditions of being: (1) just sufficient to elicit the correct response, but (2) no stronger than is required for this purpose. Regardless of the fate of this proposition in terms of further experimental verification of its implications, it clearly implies the general caution against "over-prompting" on the one hand, and "under-prompting" on the other hand, advocated by various writers on techniques of programming (e.g., Klaus, 1961). It also suggests a programming rule which, at least in principle, might be expressed in quantitative terms.

In any case, it seems clear that the most general argument for the vanishing of prompts is to prepare the learner to cope with a criterion or test situation in which the prompts will not be present. The question of vanishing thus involves an important special aspect of the use of prompting—the dual requirement that not only must prompting insure that responses be correct, but also the response context must be "appropriate." Appropriateness includes not only the additional factors discussed by Wulff and Emeson in Chapter 30, but the special condition that practice should be provided in the presence of cues that will be present and the absence of cues that will not be present in the criterion or payoff situation. The usefulness of doing this, indicated by the Angell and Lumsdaine study (Chapter 24), assumes that such explicit provision for unprompted practice should ideally be provided in the program material rather than merely depending on the learner to "ignore" the prompt when he does not need it, as may frequently be the case in fully prompted conditions such as those utilized by Cook (Chapter 21) and by Kopstein and Roshal (Chapter 20).

Fuller validation of the vanishing principle involves not only showing an advantage for partial versus full prompting, but also the question of where in the instructional sequence the more and less complete forms of prompting should be used.

The question of where overt unprompted practice should be required is, as noted by McGuire (Chapter 27, p. 6), not one that is wholly novel to questions raised in the context of the current concern with instructional programming; it is partly anticipated in studies in the experimental laboratory preceding the instructional-media experiments reported in this volume. Thus W. C. F. Krueger (1930) found that a single (unprompted) practice session should be introduced late rather than early in an instructional session. The relation of this finding to the concept of vanishing is, of course, that prompted rather than unprompted practice (whether implicit, or explicit, and/or overt) is required early in training to maximize correct response, whereas unprompted practice needs to be introduced at the point where the student can perform correctly under unprompted conditions. More systematic experimental variation in the temporal placement and "gradients" of prompting remains an important field for investigation in the design of programmed instruction.

The determination of the most appropriate gradient of vanishing for any learning situation would appear to interact with task characteristics and with the kind of terminal performance that is actually desired. For the primary case where a relatively small amount of material is being learned, and ability to meet criterion performance very shortly after the conclusion of instruction is what is required, it would appear that the vanishing gradient proposed by Lumsdaine (1959b)—see above—should be appropriate; that is, just sufficient prompting at each point to permit correct responding but also at each point no more than this amount. However, the desired criterion and conditions of learning are often not this simple. In many subject matters what is required, realistically, is recall or ability to apply what has been learned at some time considerably after the conclusion of instruction. Also to be reckoned with is the further complication that what is learned at a particular point in an instructional sequence may be overlaid with and complicated by the subsequent learning of related material which may not only extend but also interfere with what has been learned at the earlier point. Thus the principle of vanishing of cues relates inevitably to other aspects of the organization of material. An extension of the vanishing-gradient notion to the placement of review sequences to insure longer-term retention is discussed later in this connection.

Perceptual Factors in Transfer from Instruction to Terminal Performance

Any application of the general notion of vanishing must always involve a decision for reconciling the competing requirements of: (a) insuring correct response at any point in an instructional sequence with those of (b) preparing the student for later, unsupported, unprompted practice. A further aspect of the need to develop training procedures to foster transfer to later application situations is discussed by Sheffield (Chapter 5) for the case where there is the requirement of integrating component learned acts into a total task. Although this is perhaps most obviouslv seen in the case of the learning of a serial task, a counterpart is evidently to be found in the situation of many forms of conceptual learning where not only must unit skills be mastered but these must then be properly integrated into a total performance. In non-serial tasks, however, there is the special requirement (cf. also Sheffield's comments in Chapters 2 and 8 on "perceptual blueprinting") that the order in which the component response patterns may be required in a particular application is not that of a given linear sequence that has been previously practiced as such; rather, the individual may be required to "call out" the desired response components in a variety of orders, as appropriate to the particular situation. For example, in learning a language the student must combine various elementary syntactical and grammatical forms to compose a sentence, and in mathematical reasoning he must utilize any of several theorems or techniques as appropriate to the solution of a particular problem.

The perceptual-blueprinting notion as elaborated by Sheffield in Chapter 2 would appear to bear an interesting relation to the concept of Tolman (1948) of cognitive maps. Regardless of the mechanism whereby such perceptual blueprints are mediated, there can be little doubt that something like the function they are presumed to serve occurs as a useful and fairly pervasive aid to performance, and that the study of specific instructional techniques whereby this functioning can be fostered is an important area for inquiry. The "implosion" technique studied by Sheffield and collaborators in Chapter 8 is one interesting and provocative approach that appears well worth following up, despite the fact that in the experimental demonstration reported in Chapter 8 no control was provided in terms of an alternative for the slight additional amount of time required to implement the technique. A comparison of an equivalent amount of time devoted to an alternative form of review less well calculated to implement the "perceptual blueprinting" idea would be of interest, as well as investigation of alternative techniques for implementing this concept.

The use of pictorial and verbal representation in the study by Kopstein and Roshal (Chapter 20) also bears an interesting relation to the role of perceptual factors in mediating desired forms of terminal performance. This experiment was an outgrowth of an earlier investigation by Lumsdaine (1949, 1950, 1958) in which verbal and pictorial representations for paired associates were used in both the stimulus and response positions. In the Kopstein and Roshal experiment, the use of pictorial versus verbal representation was investigated only with respect to representation of stimulus terms of the pairs to be learned. The decision to do this was partly related to the finding in Lumsdaine's experiment that, at least where the required response was by definition of a verbal character, advantage for pictorial representation was found only for the stimulus term of a pair and not for the response term. An especially interesting aspect of the results obtained by Kopstein and Roshal was found in the data for testing in the transfer situation of using a verbal stimulus term when the acquisition situation had used a pictorial stimulus term. The finding was that the presence of contextual verbal material facilitated transfer under these conditions, whereas it appeared to be a hindrance when testing as well as training used pictorial stimuli. The experimental design and analysis which generated this finding might be thought of as representing one step in an experimental attack needed for disentangling the communalities and differences in paired-associate as compared with "continuous-discourse" learning. In addition, it suggests the need for approaches that can capitalize on the advantages of a superior training procedure (here, use of pictorial stimulus terms) while utilizing a mechanism (here, verbal context) for maximizing transfer to meet the needs of a different test or application situation. It is not clear that a simple "identical-elements" theory of transfer is adequate to explain the Kopstein and Roshal results, since the added elements ("A . . . is called . . .") were common to all pairs and thus did not serve directly to differentiate one pair from another.

The interaction of perceptual factors with conditions fostering positive transfer (from practice during instruction to performance in the test or criterion situation) is of central importance in considering the implications of the study by Roshal reported in Chapter 11. It will be recalled that (in addition to opportunity for overt student participation during the filmed instruction) other variables, hypothesized to effect transfer from implicit responding during demonstration to later overt performance, centered around the similarity of visual cues in the practice and test situation. The experimental variations included not only the angle of view (so-called subjective versus objective camera angle) but also the use of motion-picture depiction, as contrasted with depiction by a series of still pictures, and the delineation of the task sequence with versus without the inclusion of the demonstrator's hands.

Results for the first two of these three factors clearly supported the general hypothesis that demonstrational teaching of a perceptual-motor act will be more effective as the demonstrational representation approaches a representation of the learner himself performing the task. A reversal of the implication of this general hypothesis in the case of the third factor (hands)—especially in the simplest of the three tasks studied—is provocative in calling attention to the importance of perceptual factors suggestive of the "perceptual blueprinting" discussed by Sheffield in Chapter 2 (cf. also Chapter 8).

The reversal also calls attention to the influence of task characteristics as modifying variables, and exemplifies the situation where the implementation of a given "version" or "treatment" may in fact alter more than one factor that is psychologically important in learning. In this case, greater fidelity in representing the hands and actions of the performer decreased the clarity with which the configuration of the objects to be manipulated could be perceived during initial demonstration, thereby reducing the potential effectiveness of prompts for the adequate occurrence of implicit perceptual responses. This fact illustrates a general trade-off problem in designing effective instruction (paralleling the trade-off required between providing sufficient prompts to guarantee correct performance versus providing sufficient practice with only those cues that will be present in the test situation). The trade-off involved here is that similarity to the perceptual cues with which the student will be confronted when he attempts to practice may, if maximized, introduce the deleterious factor of reduced clarity of perception of the task. Thus, in the Roshal study, the presence of the hands could obscure the perception of the configurational sequence of the rope in which the knot is being tied. Many similar instances in the manipulation of objects would evidently be encountered in demonstrational teaching.

An interesting possibility here would be to utilize "ghost-view" photography, in which the cues of the demonstrator's hands could be perceptually present so as to mediate transfer to his own later manipulation of the task objects, but with transparent rendition of the hands so as to permit clear perception of the underlying manipulanda. This technique has been used in still photographic representations, though evidently without supporting experimentation, in providing instructions on keyboard equipment such as calculating machines. The technique suggested is, of course, peculiar to the question of demonstrational learning of those tasks where the realistic depiction of the learner or demonstrator performing the task obscures the perception of other task elements. It is evident that at least in demonstrating some tasks, lack of realism in a demonstrational sequence is essential—for example, where the ultimate task-performance criterion is that of being able to perform the task, such as disassembling and cleaning a rifle in the dark, or performing manipulations on the interior of a piece of equipment by "feel." It would be interesting to apply the principle of vanishing here in a very literal sense by progressively darkening the visual demonstration so as gradually to approximate, with increasingly minimal cues, the required performance conditions.

The frequently competing requirements between provision of adequate prompting for imitative or other responses by the student, on the one hand, and, on the other hand, similarity to stimulus condition to be encountered in the performance situation also appears in the special case of the use of slow motion, which may provide perceptually clearer cues, but at the same time provides cues which are dissimilar to those that will be encountered in the actual task. For these special cases as well as more generally, we need theoretical as well as experimental development to cope with the questions of optimum progression in the elimination of non-task-relevant cues, and of the selection of prompts of a kind which will optimally foster initial correct practice but, at the same time, will minimize interferences due to the artificial conditions which they set up.

Factors Related to the Nature and Role of the Student's Response

Under this heading we need to consider not only the question of response form but also the consequences of the student's response in terms of confirmation-feedback or reinforcement, and the consequences of explicit student response (overt or covert) in fostering generalized or "motivational" effects, as distinguished from sheer practice effects.

Explicit and Implicit Response

The status of the comparisons of overt and covert response by Michael and Maccoby (Chapter 18), McGuire (Chapter 27) and Kanner and Sulzer (Chapter 28) can be clarified by first making a distinction between "explicit" and "implicit" responding. It seems desirable to clarify this distinction because the term "implicit response" has frequently been used to embrace both explicit and implicit covert responses in the sense here defined. For the purpose of this discussion, an explicit response means a response given by the student on some specific occasion identified to him through instructions—e.g., the leaving of a blank which he must fill in or a question he is to answer. Explicit responses may be either overt or covert. Overt responses are, of course, those which are externally observable, such as saying or calling out the response, writing it, marking a worksheet (cf.

Kimble and Wulff, Chapter 15, for example), spelling out a response on a keyboard, or pushing a multiple-choice selector button. An explicit covert response, on the other hand, generally means merely "thinking" the response or saying it to oneself in response to a specific cue but without leaving a record or yielding clearly distinguishable external evidences. The further distinction to be made is that between explicit covert responses of this kind and implicit responses of a covert sort. The latter are responses for which the occasion is less clearly identified, such as subvocal or other responses that occur in silent reading, in listening to a presentation, or in creative thinking. Although we may identify some general occasion or stimulus for such implicit responses, their defining characteristic is that they are not only covert but are not explicitly structured in response to specific experimenter-provided or program-provided stimuli. (However, they may occasionally have fragmentary overt manifestations, or can under special circumstances be translated into overt responses, as in Newell, Shaw, and Simon's (1958) attempts to get overt records of the usually implicit train of covert responses that an individual makes in thinking about the solution of a problem.)

Theorization about implicit student responding is relevant to several of the studies in this volume that concern the structuring of stimulus materials designed to foster such responding—e.g., the studies by Wulff and Kraeling (Chapter 10), by Roshal (Chapter 11), and by Lumsdaine, Sulzer, and Kopstein (Chapter 17). However, specific experimental comparisons between the functions and consequences of overt and covert responding tend naturally to concentrate on explicit responses. A partial exception is found in some of the recent studies of programmed instruction (Goldbeck, 1960; Goldbeck, Campbell, and Llewellyn, 1960; Roe, *et al.*, 1960 a, b; Holland, 1960) which have compared explicit overt or covert responses, in the form of filling in blanks in an instructional program, as against the implicit-response condition of simply reading the corresponding material with the blanks already filled in. A borderline situation occurs in those cases (e.g., Goldbeck, 1960) where the words which are read (and which in the overt-response condition would be filled in by the students as explicit responses) are underlined or otherwise stressed in the comparison condition. This presumably would tend to foster conscious, explicit making of corresponding covert responses in the context of a less structured stream of implicit responding that would occur during the course of undifferentiated silent reading.

Experimental Evidence on Overt versus Covert Responding

The experiments reported by Michael and Maccoby (Chapter 18) and by Kanner and Sulzer (Chapter 28), as well as several subsequent experiments conducted in the context of auto-instructional devices, have in general come out with the result of failing to find a difference between the effectiveness of the two forms of responses. Such a null outcome has the usual difficulties of interpretation and would generally seem to be less useful after the fact than such results as those of McGuire in which, under certain conditions which were partially predicted on theoretical bases, one or the other form of response (in this case overt) showed an actual disadvantage under specified conditions (rapid rate of presentation). However, the lack of differences reported by Michael and Maccoby and by Kanner and

Sulzer between overt and covert responding, both for "continuous-discourse" material in the one case and for paired-associates material in the other, may be considered to be of some interest in the light of two characteristics of these experiments: (1) the fact that the number of cases and other aspects of sensitivity were such that any rather substantial differences, if present, would probably have been revealed as significant; and (2) that the comparison was made in the context of co-variation with several other factors, including presence and absence of feedback to the learner, rather than only in a single experimental context. By contrast, relatively small and insensitive experiments, such as those reported by Evans, Glaser, and Homme (1959) provide less basis for forming conclusions about the importance of the experimental variable unless they come out with positive results.

Cook and Spitzer (1960) have also concluded that, for some instructional situations, covert or mental responses were apparently about as effective as overt responses, and Goldbeck (1960) has concluded that merely reading brief passages in which key words are underlined may not be significantly less effective than having students make the active, explicit response of filling in blanks mentally or in writing. Similar results were obtained by Roe et al. (1960a). However, a study reported by Holland (1960) suggests that whether an effective contribution for overt responding is found may depend on the relevance of the responses as well as how well they are cued. For his subjects, overt response tended to be more effective, but only for when the responses were selected as the most relevant ones and were also well cued to insure their being correctly evoked.

An important variable differentiating some later experiments from those reported here by Michael and Maccoby and by Kanner and Sulzer is the factor of time. Unlike some later investigations using self-instructional materials (e.g., Goldbeck, 1960; Evans, Glaser, and Homme, 1959), the time allowed for overt and covert responding was here, in the context of classroom film instruction, the same for both forms of response; thus the present studies (Chapters 18 and 28) did not encounter the methodological problem faced by later investigators of overt-covert comparisons, in which the lack of superiority for overt response had to be evaluated in the light of the further fact that it took more time than covert responding. The overt-covert variable was also one of the two factors introduced in the further experimental analysis reported by McGuire in Chapter 27. In this case time was also manipulated as a variable, but it was not confounded with the overt-covert factor. Rather, the two factors of response mode and time were independently co-varied, and a comparison was made between overt and covert responding under fast versus slow rates of presentation.

A distinction between theoretical and practical interests arises from the different ways of treating time in later auto-instructional experiments as compared with the Chapter 18 and 28 experiments. Although, for experimental purposes, it seems clearly desirable to study the effect of overt versus covert responding without confounding differences in time required for response, it is also clear that one of the advantages of covert responding, to the extent that it is appropriate, is that—particularly as compared with laboriously writing out of the lengthier responses required in many Skinner-type self-instructional programs—the saving of time effected by covert responding is in itself an important dependent variable. Indeed, the use of time as a principal dependent variable is of crucial concern in educational measurement, and particular attention is focused upon it by the fact that various characteristics of self-paced auto-instructional programs (including the special case of overt versus covert responding) generally do require markedly different amounts of time. For example, a dilemma is encountered in interpreting the results of Coulson and Silberman (1959) on multiple-choice versus constructive responding where, if time is taken into account as a dependent variable, the interpretation of results may be quite different from the conclusions that would otherwise be drawn.

Practice Versus Motivational and Other Generalized Effects

One of the questions which arose from the phonetic-alphabet experiment of Hovland, Lumsdaine, and Sheffield (1949) was the resolution of the effects of student-response procedures in terms of the direct effects of practice, versus motivational effects that would show up in better attention and hence better learning on material not specifically practiced. The evidence they presented suggested that both kinds of effects would operate, in that the superiority of active-review over passive-review groups was diminished when added external motivation was provided (see Chapter 1). However, this analysis did not isolate practice effects from motivational effects, as was done in the two experiments reported in the present volume by Michael and Maccoby (Chapter 18) and by Kanner and Sulzer (Chapter 28). Both experiments employed the procedure of giving some experimental groups practice, through active (explicit) response procedures, on only half of the material to be learned. It will be recalled that both for substance learning (Chapter 18 and first section of Chapter 19) and for rote associations (Chapter 28) the results showed marked gains on the material expressly practiced but no significant gains on the other half of the material. These results of course point to practice as the key factor. However, both common sense and the earlier results of Hovland, Lumsdaine, and Sheffield (1949) would suggest that motivational carry-over effects could also be expected at least under conditions of generally low extrinsic motivation. It was these considerations that led to the further experiment, by Maccoby, Michael, and Levine (second section of Chapter 19), which deliberately used a rather uninteresting subject matter. In this case the results showed gains both on the practiced and the non-practiced material, but only for groups that did not have the extraneous incentive of being told they would be tested. The gains on attributed motivational factors were, however, at best smaller than the direct practice effects. These results not only emphasize the importance of practice specifically, but also call attention to the need for parametric studies which can lead to dependable estimates of the relative magnitude of effects likely to be associated with different variables, rather than merely inquiring whether variable X does or does not make a significant contribution to learning. They further illustrate the general proposition that the development of a science of programmed instruction needs to seek contingent generalizations in which the importance of one factor is predicted in relation to the influence of other modifying factors rather than in absolute terms.

Clearly to be distinguished from transfer effects of a "motivational" sort presumably related to general enhancement of attention (Chapters 18, 19, and 28) are the kinds of generalization effects demonstrated by Kimble and Wulff in Chapter 15, in showing that guidance of practice on illustrative examples in the skill of slide-rule scale reading was effective in improving performance not only on the materials specifically practiced but also in dealing with similar examples involving the same skills. Here there is an explicit basis for transfer in that techniques or skills of an inherently generalizable sort are being learned, as compared with the case of discrete material, whether continuous-discourse or paired-associate (as in Chapters 18 and 28), in which one would expect carry-over of the effects of practice procedures only through generalized attention or motivational enhancement. In the Chapter 15 study, the direct transfer or generalization effects thus represented a dependent variable of primary interest, whereas, in the concern of Chapter 18 with generalized motivational carry-over, it was necessary to take special precautions to control against contaminating the data by possible generalization from practiced material, by studying the gains on discrete, unrelated material to which direct transfer would not be expected to occur.

Feedback, Knowledge of Results and Reinforcement

The function of feedback or correction/confirmation following the student's response is closely related to the role that the response itself is conceived to have, and also to the degree of prompting prior to responding. A distinction has previously been made between two different functions of a confirmation/correction panel in a teaching-machine program. The first function is that of reinforcement in the sense of reward, presumed to strengthen the benefits of immediately preceding responding. The second function is that of providing prompts or cues for an additional implicit response, which may supercede, repeat, or fill in, at a covert level, the immediately preceding response, depending on whether that response was, respectively, in error, correct, or lacking.

The studies of Michael and Maccoby (Chapter 18) and Kanner and Sulzer (Chapter 28) seem clearly to demonstrate that, whatever function is served, providing knowledge of the correct response (KCR) can be an important factor in the effectiveness of learning. However, the feedback function may be seen as less important, if one pits such KCR (which, aside from its possible reward function, clearly provides prompting for an implicit correcting or confirming response) against the provision of adequate prior prompting to insure correct response in the first place, as was done by Cook and collaborators (Chapter 21). Suggestive anecdotal evidence along these lines is furnished by the frequent observation that, with a well-prompted program of the Skinner type, students may frequently pay little attention to the confirmation/correction panel since the program is sufficiently well cued so that they are generally certain of the correctness of the response they have made. In any case, with better prompting, the importance of the confirmation/correction function (and, therefore, of variations in it) would appear to be less. A recent study by Angell and Lumsdaine (1961a) bears out this supposition (see Appendix). They found that, in paired-associate learning without prior prompting trials, variations in the completeness of feedback information made a considerable difference, but that the importance of these variations was slight when sufficient prior prompting had been given. The Chapter 18 and Chapter 28 results on the importance of "KCR," as well as the conclusions earlier drawn by Hirsch (1952) —cf. also Briggs (Chapter 23) and Stephens (1953)—concerning the superiority of more complete information in giving feedback after a response, would appear to require reinterpretation in the light of these latter findings.

In addition, the relative importance of the informational versus the reward functions of feedback might be expected to change radically with variation in the strength of initial prompting. When prompting is sufficient to reduce the informational function of feedback, its reward function would be expected to interact with and depend on general motivational level and other motivational factors in the situation. For example, the reward effect would be expected to be greater with lengthy programs than with short ones (where the problem of boredom is not so likely to arise), and to be greater with a relatively uninteresting program than with one which is inherently interesting because of compelling subject matter or sprightly writing. This implication parallels that from the results of Chapter 19, which indicated that the motivational effects attributable to participation exercises—including, of course, the reinforcing effects of response correction—were greater under low-motivation conditions.

In addition, the confirming or reward function of feedback would be expected to be greater when we are concerned with response-shaping requirements that are not typical in many instances of academic learning and technical training. A further discussion of this point is given elsewhere by the writer (Lumsdaine, 1962b). In brief, it can be contended that concentration on the manipulation of reinforcement contingencies is primarily of use only in those instances where the experimenter or instructor has, at the outset, no ready way of eliciting a desired ("operant") response. This would appear to be the case in a wide range of important instructional objectives, such as (to name a few) the articulation of new phonemes in second-language learning, the acquisition of proper postural and motor-coordination responses in some kinds of musical and athletic skills, and the effective development of various forms of socially important behavior. For a great deal of academic learning, however, the need for manipulating reinforcement contingencies to shape a desired response can be largely by-passed because the responses themselves have already been learned and can be readily elicited by instruction or demonstration. Here the principal problems would appear to be those of transfer of stimulus control for responses, verbal and symbolic, which the learner has already acquired or for which learning to make an adequate response poses no particular difficulty. (See also Skinner, 1961.) In this case, cue manipulation seems to be the critical variable, and theory based on association of cue and response by temporal contiguity seems more appropriate than theories that depend on the retroactive influence of reinforcement (cf. Zeaman, 1959; Lumsdaine, 1962b; and the theoretical development by Sheffield in Chapter 2).

Finally, even where response shaping is required, the possibilities deserve exploration (with human subjects) for using such techniques as demonstration-imitation, and perhaps even direct motor manipulation in some cases, to short-cut the process of merely waiting until some approximation of the desired terminal behavior occurs and then reinforcing it. Thus there appears to be an open question for many forms of motor activity—such as handwriting, musical and athletic skills—as to how much the instructor must rely on merely reinforcing correct behavior when it is spontaneously emitted, and to what extent he can proceed more effectively by exploiting techniques through which the desired behavior or an approximation to it can be elicited more directly through the manipulation of demonstrational and other forms of guiding stimuli.

Some Factors in the Organization and Sequencing of Instruction

The organization or sequencing of the elements of an instructional program, as well as its pacing, clearly affect the probability of correct response and thus will be expected to interact with the use of prompting techniques within a frame and with related factors that affect the probability of correct response. Nevertheless, it appears convenient to distinguish organizational factors that are descriptive of the total organization of a program, or major segments of it, from the factors of response-cueing that operate most evidently at the molecular level within a particular self-instructional frame or segment. Some of the more molar questions of organization dealt with in this volume concern the sequencing and grouping of material, massing or spacing of similar and dissimilar elements, and organizational structure in relation to supra-ordinate communalites or "context cues." Superficially at least, there appear to be two distinguishable aspects to the question of organization: (1) organizational features intrinsic to the task to be learned, and (2) organization of instruction through extrinsic measures superimposed on the inherent features of the task. The clearest case of the latter might be formed in attempts to structure undifferentiated lists of paired associates or rote serial sequences by grouping them arbitrarily into differentiated sub-sequences (cf. Chapter 26). In several of the studies reported herein, however, the procedures used in the sequencing or structuring of instructional elements were either chosen in advance or analyzed after the fact in relation to the organizable features of the task itself (e.g., Chapters 6-9 and 10 for serial-procedure learning, and Chapters 29, 30 for paired-associate learning). An intermediate case lies in the initial basis for choice of practice-and-demonstration sequences in Chapter 3, where the lengths of practice segments were in part chosen empirically in relation to demonstration-assimilation-span data rather than wholly with respect to structural features of the task.

Relations of Instructional Sequence to Task-Organization Features

The factor of spacing of demonstrational and practice segments, or length of demonstration segments before practice (when total amount of practice and of presentation is held constant), is central in the studies reported in Chapters 3, 4, and 5, by Maccoby, Sheffield, and collaborators. Considered at one level of discourse this is an aspect of the whole-part problem, but it can also be related conceptually to the effective strength of prompts at the points when the student is asked to practice. The whole-part organizational question, dealt with in Part I for the case of procedural learning (but having counterparts for other forms of instructional objectives), also involves further questions concerning the integration and consolidation of part-learnings. Here such further problems are posed as the decision between "repetitive" versus "consecutive" patterns of organization, studied by Margolius, Sheffield, and Maccoby in the experiments reported in Chapter 6, and the timing of demonstration and overt practice as a function of task organization into "natural units" (Chapter 7). With respect to the latter it is evident that more experimentation is needed, especially since the hypothesis under consideration could not be given a clear-cut test in these experiments in view of the high degree of inherent organization of the task that was used. Of interest in this connection are the competing requirements for consolidation of sub-task learning (by practice at the end of an inherently organized sub-unit) versus serial integration of the several sub-units fostered by practice not thus synchronized. This is one of the instances where theoretical and experimental analysis can be facilitated by, and reciprocally may contribute to, an improved taxonomy for classifying tasks in terms of their structural features (cf. also Wulff and Emeson, Chapter 30).

The theorization developed by Sheffield in Chapter 2 is instructive as an example of applying theoretical conceptions that are applicable at a wide level of generality but with modification and extension in relation to the task features of a specific category of learning tasks. This theorization, developed from a substructure based on perceptual cross-conditioning mechanisms and on conditions of stimulus similarity that should foster transfer from acquisition to performance, provided an integrated theoretical sub-system that takes into account identifiable task characteristics in serial-learning situations, including the structuring of the task into so-called "natural units." It will be recalled that these were defined as units which tend to hang together perceptually because of a common context-cue matrix for each such natural unit, which can serve to reduce intra-serial interference between component responses required within one natural unit and those required in a separate unit with a different cue-context. Of particular interest is the tentative contingent generalization that—since arrangement of practice and demonstration segments can affect both the reduction of intra-serial interference and the integration of sub-sequences into total task performance—the method of choice with respect to both repetitive-versus-consecutive practice and synchrony-asynchrony of practice with natural units will differ depending on the degree of inherent task organization.

At the same time, the potential parallelism in the functioning of organizational variables across classes of tasks that are superficially quite different is not to be discounted. As Sheffield and Maccoby remark (Chapter 9), "The principles involved are much the same as those which apply to the difference between a poorly organized lecture and a well-organized lecture." The extent to which devices for structuring academic subject matter can be expected to contribute to learning would, as in the case of the procedural tasks dealt with in Part I, depend on the degree of inherent structuring of the subject matter. This consideration might in part explain the difference between the negative findings on the value of organizing subtitles obtained by Miller and Levine (1952; cf. Appendix) and the later positive findings reported by Northrop (1952) for the advantage of using subtitles and related devices to enhance the organizational structure of films.

A further interesting parallel suggests itself between the considerations underlying consecutive versus repetitive organization for sequential tasks (cf. Chapter 7) with the recently proposed "spiral" concept of organization suggested by Glaser (1962) for the programming of conceptual materials. At least some of the problems in terms of integration versus consolidation would appear to be fundamentally similar, and one might suggest that the spiral form of organization in which successive "cuts" through the total subject matter are taken, rather than mastering each unit as it comes, would tend to be relatively more advantageous for instructional subject matters in which the requirements of integration predominate rather than those of internal consolidation.

One other possible parallel across dissimilar classes of learning tasks might be suggested in considering the role of so-called "context cues" and of "class-descriptive cues" in their functioning as prompts to avoid interference from competing responses in two different kinds of tasks studied by contributors to this volume. The first instance involves the context-cue rationale utilized by Sheffield, Maccoby, and collaborators for procedural tasks in Part I; the second is found in the use of class cues to foster discrimination, in somewhat analogous fashion, in the case of the paired-associate learning task studied by Wulff and Stolurow (Chapter 29). In the former case, the essential rationale is that individual elements, if practiced within a distinctive context of their own sub-task, will tend to be elicited only in that context and therefore will not compete with similar elements in other sub-tasks. In the case of the Wulff and Stolurow study, analogously, distinctive discriminative responses appear to be more easily fostered when competing response tendencies with respect to a particular item characteristic are differentiated by being brought together in a common "context" or frame of reference. The analogy is obviously not a perfect one, however, and it is not necessarily to be expected that a single general principle will apply to both cases.

Other Aspects of Sequencing and Organization

It will be recalled that the temporal placement of functionally different but complementary instructional procedures also entered into the study reported by Wulff and Kraeling (Chapter 10) in the comparison of familiarization procedures placed immediately prior to each step of a procedural assembly, versus earlier familiarization given before the commencement of assembly training. Parallel problems in the placement of the familiarization procedures for developing relevant perceptual discriminations would occur in the organization of verbal-conceptual learning as well as in procedural training. For example, should definitions and terms be learned as a group in advance of instruction on the principles into which they enter, or should each be introduced as it becomes appropriate? This problem is similar to the controversial and experimentally unresolved question of temporal order and contextual integration in the learning of lexical elements of a language in relation to experience in their use in syntactical patterns. The potential conflict between factors favorable to task integration and those favoring perceptual discrimination has been noted earlier in discussing the cueing of implicit responses as exemplified in the Wulff and Kraeling study.

A further problem in the organization and selection of content for instruction that relates to preparation for subsequent criterion performance is the range of illustrative examples that are used in the instruction. One of the questions that may be distinguished is that of the extent to which varied rather than identical repetition of illustrations should be employed. There would seem to be little question that, at least for some stages of learning and degree of task difficulty, repetition of identical examples may be preferable from the standpoint of stimulus constancy to rapid shifting from one example to another. On the other hand, clearly the use of varied examples should promote the student's ability to generalize. The inconsistency of the results reported by Kanner and McClure (1956; cf. Appendix) suggests that the identification of conditions under which one or the other form of repetition is preferable remains a task for the future. The results obtained by Rimland (1955) in teaching knot-tying, using as instructional materials several of the film versions devised for the experiment by Roshal (Chapter 11), have some bearing on the question. Rimland reported that the repeated use of a single form of demonstration was preferable to the use of varied forms of the demonstration. However, it is not clear to what extent this result is attributable to the fact that the different forms of demonstration were, as shown by Roshal, of differing effectiveness (when each was used alone). In the Kanner and McClure study, on the other hand, the variety was merely in the specific content of the illustrative examples, and the pattern of presentation for the different examples was constant. The use of reiteration of material in identical form versus use of varied examples is also related to McGuire's notion (Chapter 13) that the teaching effectiveness of a short silent film can be improved by the addition of a sound track which provides "simultaneous repetition" in a different modality instead of doubling the length of the film, which provides successive repetition in the same modality.

Modifying factors that would seem pertinent in assessing the relative contribution of varied versus identical repetition include task difficulty and stage of practice. With difficult illustrative material, one would certainly expect instances in which identical repetition of the same example (perhaps then followed by varied presentation of that same example) should precede the introduction of further examples involving the same procedures or principles but differing with respect to the specific illustrative elements. General problems illustrated here are the needs to determine gradients of generalization and discrimination that are involved in various classes of tasks, and to determine the extent to which, as contrasted with empirical determination anew for each subject matter, it is possible to formulate (and validate experimentally) general rules relating task characteristics to the manner in which alternative patterns of repetition should be employed.

One additional problem in the organization of academic subject matter that is deserving of comment in relation to the repetitive-consecutive arrangement for segments of serial procedures studied in Chapter 6 is the placement of review sequences in academic subject matter. Miller and Levine (1952) and Miller, Levine, and Sternberg (1956)—see Appendix—compared two different ways of placing such sequences. Two subject matters—Ohm's Law and defense atomic attack—were used in replicative experiments for comparing "massed" review sequences versus "spaced" review sequences. In the spaced-review procedure, each section of the film was reviewed immediately after its initial presentation (analagous to the "Repetitive" condition for the serial task studied by Margolius, Sheffield, and Maccoby in Chapter 6). In the massed-review procedure, all of the main presentation sections of the film were shown before any review, and then the reviews of each of the sections were grouped together at the end (analagous to the "Consecutive" condition in Chapter 6—where, however, the "review" was identical to the first presentation instead of being an abbreviated recapitulation as in the studies by Miller et al.). The results obtained by Miller and collaborators showed massed review to be superior to spaced review. One partial explanation invokes the fact that "massed review" amounts to spaced practice of each section (since the review of each section is separated in time from the initial presentation by virtue of the fact that all of the reviews are deferred until the end of the lesson). An analogy with the differential prediction made by Margolius, Maccoby, and Sheffield for serial learning might predict that the relative advantage of the two review-spacing arrangements would depend on subject-matter characteristics such as the degree of integration or common context within the several sections of the topic. It is not clear, however, that this analogy would hold, in view of the difference in terminal behavior requirements between serial performance and substance learning in which serial-integration factors are less important.

In substance learning, the optimum form of organization of review could well change when we consider educational criteria other than immediate recall. For example, if retention over a considerable period of time is required, it could be hypothesized that successive reviews of the same material with intervening instruction on other topics should be spaced at increasing intervals, so as to approximate more and more nearly the conditions of long-term retention after the interposition of interfering experiences. This hypothesis is considered in more detail elsewhere (Lumsdaine, 1962b) as an extension of the principle of vanishing; the prediction is, in effect, that the introduction of successive reviews should occur just at those points where the learner can barely perform correctly but, for maximum efficiency, should not be more closely spaced than this.

A still further elaboration is suggested if we consider that the realistic objective of education often must be to create conditions such that the learner will merely be able to relearn the desired behavior efficiently, rather than demanding long-term recall without his first "brushing up." This would suggest as an appropriate measure of retention a "savings" index (which is familiar to students of experimental psychology but has seldom been used as a dependent variable in experiments on instruction). In this case, however, the implication for spacing of reinstatement sequences in a self-instructional program might well be different from that suggested above, with the review of material deferred until the student has gone considerably past the point where he can at once perform adequately in a review sequence. This would imply a change not only in the spacing but also in the character of the review sequences, so as to provide opportunity for students to acquire proficiency in relearning material rapidly on the basis of a minimal review sequence.

Some Implications for a Science of Instruction

There are several assumptions and points of view implicit in the foregoing discussion and in the conception underlying the studies reported in this volume. Major assumptions stated at the outset of this chapter are the need for developing a "science of instruction," and the notion that experimental inquiry to this end must be guided by relevant theory, which generally should be related closely to the implicit or explicit responses of the student. A further assumption lies in a distinction which Skinner and others have made between experimental studies which merely observe the conditions under which learning and forgetting take place and those which place emphasis on the management of learning conditions designed to create a desired form of behavior. The latter orientation—which would seem to be more conducive to the development of a science of instruction—is adopted, to varying degrees, in the studies reported in this volume.

A further assumption is that the requirements of a science of instruction (or, more specifically, a science of programmed instruction) can be distinguished from that of a more basic generalized science of learning, on the one hand, and, on the other hand, from merely an educational technology in which the findings of a basic learning science are translated into practice. These distinctions are of interest in relation to Skinner's 1954 paper, "The Science of Learning and the Art of Teaching," and are more explicitly reflected in both the title and some of the content of Melton's discussion of "The Science of Learning and the Technology of Educational Methods" (1959). It would be hard to disagree with Skinner's basic emphasis on the possibility for applying in educational practice the methods developed in the laboratory investigation of behavior, or with Melton's convincing argument for the fruitfulness of two-way interaction between the development of basic learning science and of educational technology. Melton's further point that the character of basic research derives more from the investigator's purpose and approach than from the class of learning tasks investigated is especially appropriate in the context of the present studies, being perhaps best exemplified in the theoretical developments elaborated by Sheffield in Chapter 2. Particularly relevant also is his emphasis on the need to relate instructional variables to task characteristics and, to this end, to develop a better task taxonomy for learnable human performances.

It can further be contended that there is an important "middle ground" between a basic science of learning and an applied technology of educational method, and that there is both room for and need, in this middle ground, to develop a science of instruction. This argument rests in part on two assumptions. The first assumption is that the principles or rules needed to guide the effective management of learning through programmed instruction for developing a wide range of human capabilities cannot be deduced solely from more general propositions, applicable to any and all forms of learning, discovered in the psychological laboratory. The second assumption is that the formulation of such principles or rules can be generated through development of appropriate theory and experimental inquiry which relates instructional variables to specific task-parameters and learner characteristics appropriate to identifiable classes of human-learning tasks. Perhaps the most significant aspect of the studies reported herein is that they illustrate some of the possibilities and problems involved in this kind of endeavor.

Two further notions may be helpful in understanding the position that a specific science of instruction must be developed through studies of programmed instruction on a variety of tasks. The first notion involves the concept of "levels" of principles having different degrees of generality and precision in relation to learning and instruction. Three such levels were, for example, distinguished by Hovland, Lumsdaine, and Sheffield (1949) in relation to their experimental studies of the effects of mass-communication media on knowledge and belief. They distinguished (1949, p. 4): (1) basic learning principles—common to all educational devices; (2) mass communication principles—applying to films and similar communication media; and (3) film principles—related specifically to the medium of films. The writer's present contention is that the dimensions of a science of programmed instruction[3] might be delineated by an analagous though somewhat different distinction between several levels of principles or procedures. These might be identified as:

1. Basic learning principles—applying to all forms of human and animal learning (the province of a science of learning);

2. Instructional principles—applying to broad yet well-defined classes of learning objectives that are sought by instruction designed to create specific classes of human capabilities (the province of a science of instruction);

3. Programming rules and procedures—rules of thumb found empirically to work in the design, construction, "de-bugging," and use of specific programmed-instruction sequences, for a particular subject matter, to be used with a defined sub-population of learners (the province of technological application in educational practice).

The principles toward which theorization and experimental findings, developed in such studies as those herein reported, might hopefully converge are not necessarily restricted to any one of these levels of generality. As emphasized by Melton (1959) there is no necessary reason why studies conducted with "practical" learning tasks, such as school subject matter or technical or procedural skills technicians must learn, cannot be formulated and conducted in a way that would contribute to the elucidation of principles at the most general basic learning level.

The argument here, in any case, is that theoretical and experimental inquiry in the second of these categories, dealing with generalizable classes of task variables as related to learner characteristics as in some of the studies in this volume,

[3]The limitation to a science of programmed instruction does not necessarily preclude a wider science of instruction transcending programmed media, but the former is clearly the more easily realizable goal in view of the more directly reproducible nature of the events mediated by programmed-instruction sequences (cf. Chapter 1).

is necessary and ultimately can be sufficient for the development of principles appropriate to a science of instruction. One way of characterizing the domain of such a science of instruction, as contrasted both with a more general science of learning and with more specific particular technological applications, is to say that the science of instruction must deal explicitly with classes of task variables and with intervening implicit-response variables. The latter variables enter importantly into symbolic and perceptual learning—including that mediated through reading and through listening—though, as such, they generally would not enter into the formulation of basic learning principles concerned at a level of generality which applies to learning by simpler organisms. The identification of this middle area of instructional science as a separate discipline in no way implies, of course, that sharp or rigid demarcations are possible; as in other fields, the boundaries between categories are inevitably hazy and shifting ones.

The judgment as to whether an instructional science is possible can perhaps be based in part on assessment of the likelihood of useful middle-level principles emerging from further studies of the sorts reported in this volume, and this can probably best be arrived at by review of some of these studies. It is believed that such a review will afford some basis for conviction that intermediate-level principles dealing with instruction can eventually be developed which, on the one hand, will permit considerably more precise prediction and attainment of instructional outcomes than is possible solely on the basis of universally applicable basic learning principles and, on the other hand, can achieve sufficient generality within definable sub-classes of instructional situations to transcend mere ad hoc "task engineering" (cf. Wulff and Emeson, Chapter 30).

The need for (and probable character of) such middle-level principles on which to base a science of instruction can perhaps be recapitulated in terms of two basic propositions. The first proposition is that, in view of the complexity of human learning, we can reasonably expect to find few universal generalizations that would hold for all classes of instructional objectives, all classes of learners, and all conditions of instruction. Rather, it seems evident that what is needed to account maximally for variance in the effectiveness of instruction is a series of contingent generalizations which take account of the interactions of variables so as to specify the effects of one variable in terms of other variables. Primitive illustrations in the present volume are to be found in such contingent propositions as those concerning the motivational effects of practice (Chapter 19), the value of overt practice as related to pacing of instruction (e.g., Chapter 27), or the relation of needed repetition to task difficulty and intelligence (Chapter 17).

The second proposition is that a crucial class of modifying variables that need to enter into such contingent generalizations comprise identifiable task characteristics that determine very specifically "what is to be learned" (cf. Chapter 30) and that are taken into account in determining the method of choice with respect to cue, response, and organizational variables (e.g., Chapters 6, 7, and 9). As Melton (1959) has pointed out, a needed step in this development is the evolution of an appropriate taxonomy of task characteristics. That such a taxonomy will not necessarily follow the lines of cleavage between rote verbal learning, connected-discourse learning and perceptual-motor learning has been suggested above in terms of parallels identified between the operation of variables in the several parts of this volume.

With respect to experimental investigation, the above-emphasized stress on the need to seek contingent rather than absolute generalizations implies the frequent use of factorially designed experiments in which the operation of one variable can be observed as it interacts with other variables. However, this does not argue for large "shotgun" multi-factor experiments. Rather, the optimum strategy would seem to be that of successive phases of research in which relatively simple experiments are conducted so that the implementation and, indeed, the definition of variables in later experiments can build on "plateaus" established or suggested by the outcomes of prior experimentation. The conduct of each experiment is likely to result in insights that will modify the way in which it appears most fruitful to define and implement variables in subsequent experiments. Thus, one of the most important products of earlier experiments may be to yield a progressively better idea of how to ask the right experimental question. This clarification is likely to result not only from the results of any experiment but also from experience in theorization about and implementation of the experimental variants to be compared. At least some limited progress in this direction may be perceived in the progression of studies on the role of student response reported herein, stemming from the study of an "audience-participation" procedure in the initial experiment by Hovland, Lumsdaine, and Sheffield.

In addition, the attempt to develop instructional materials that not only provide a vehicle for experimental tests of hypotheses but also seek to attain a high degree of effectiveness tends to generate provisional "rules" for effective instruction which are implemented in the experimental materials, even though they are not tested directly within the confines of the experiment. Such provisional rules, particularly when predicated on theoretical considerations, afford an important source of hypotheses for further experiments. Some good examples can be found in the working principles for constructing procedural demonstrations outlined by Sheffield and Maccoby in Chapter 9 (cf. also the somewhat similar rules proposed by May, 1946, 1958). One point of departure for useful further research would be to subject some of these propositions to experimental test, as has been done in part for analagous rules, formulated for persuasive communication by Hovland, Lumsdaine, and Sheffield (1949, pp. 203-205) in later studies reported by Hovland, *et al*. (1957). Similarly, it may be hoped that proposed rules for the programming of academic subject matter, generated as working guidelines in technologically oriented program-development efforts, will increasingly be translated into hypotheses that can be subjected to experimental tests calculated to contribute to the evolution of a science of instruction.

APPENDIX

Supplementary Research Notes

Introduction

The purpose of these research notes is to make available summaries of a number of papers, mostly unpublished, which for various reasons could not be included within the text of the present volume, but which relate closely to one or more of the volume's chapters. No attempt has been made to seek consistency in length or detail of reporting, which has been governed by editorial judgment based on the relevance and accessibility of the data reported. As noted in the preface, these summaries were drafted by Yasuko Filby, though in some cases the drafts have been condensed, expanded, or otherwise modified by the editor. The original reports of each study summarized were consulted in detail in preparing the summaries; however, a useful contribution was also made by a collection of research abstracts of these and other studies earlier prepared by Dr. John O. Cook at the AFPTRC Training Aids Research Laboratory. The research summaries are arranged alphabetically by author. A few comments indicating the genesis and interrelations of some of the studies are given below. The reader is referred to TM&PL (Lumsdaine and Glaser, 1960), pp. 574-579, for information concerning abbreviations and the availability-status for various categories of Air Force reports.

The two studies reported by Angell and Lumsdaine (1961a, c) and the study reported by Guthrie and Lumsdaine (1961) were, together with the experiment reported in Chapter 24, conducted by the American Institute for Research as part of an investigation of cueing techniques being carried out under Contract No. AF 49(638)-681 for the Air Force Office of Scientific Research (AFOSR). A related Air Force study on prompting and confirmation has been reported by Sidowski, Kopstein, and Shillestad (1961).

The four studies by Miller and Levine (1952), Miller, Levine, and Sternberg (1954), Miller, Levine and Sternberg (1956), Miller and Klier (1956b) comprise a series of experiments, conducted at Queens College under Air Force Contract No. AF 18(600)-38, in which the effect of placement of review was studied. In the first two of these studies, the "structuring" device of using subtitles to demarcate major and minor sections of the film was also varied experimentally. The second study was designed to replicate the first one, with the same experimental materials but using a military population instead of high school subjects. The third study, which again used high school subjects, made a similar experimental comparison on review placement (spaced versus massed) using a different film subject, and the fourth study was an attempt to compare spacing of overt-response reviews as well as the "passive" reviews used in the preceding three studies. Other studies involving variation in the use of repetition or review include those by Fairbanks, Guttman, and Miron (1957) and by Kanner and McClure (1956). The former study explored possibilities for introducing additional repetition within a given time period by use of "speeded speech" technique; the latter compared repetitions of identical examples with repetition using varied examples in teaching micrometer reading.

Other related studies summarized here are those reported by Kimble (1955b) and by Miller and Klier (1956a) on the effect of interpolating quizzes between instructional sessions. Both studies used military populations; the instructional material for the first dealt with the use of the slide rule, and the second dealt with principles of lubrication. In a related study by Kimble (1955a) the activity interpolated between two instructional sequences (on different aspects of the use of the slide rule) included praising or reproving students for their performance on the quiz covering the first part of the instruction. The two studies by Kimble were conducted as adjuncts to the experimental work reported in Chapter 15. The effect of brief rest-pauses interpolated between sections of an instructional film is reported by Faison, Rosé, and Podell (1955a), who employed infra-red cinematography to obtain indices of audience behavior during the showing.

The studies by Wulff (1955) and by Weiss and Fine (1955) both deal with the effect of prior name-learning upon learning. The first of these studies is concerned with the effect of prior nomenclature training upon the learning of a mechanical assembly task (cf. also Saltz and Newman, 1960), while the second deals with substance learning (structure and functions of the United Nations). Also concerned with the effects of preparatory procedures preceding instruction is the study by Romney (1952), in which the concern was with the effect of "set-inducing" instructions.

The studies by Gropper, Lumsdaine, and Shipman (1961) and Zuckerman (1951, 1954) were both concerned with the technological application of test data on learning from a preliminary version of an instructional presentation for predicting or improving the effectiveness of a revised presentation. The former was limited to prediction of the effects of a polished-up film presentation using the same narration as the preliminary version, while the latter study carried through with use of the student response data to revise and improve the presentation.

- - - - - - - - - - -

Angell, D., & Lumsdaine, A. A. The effects of prompting trials and partial-correction procedures on learning by anticipation. San Mateo, Calif.: American Institute for Research, September 1961 (a). Report No. AFOSR-TN-61-1343.

Two experiments, performed under an AFOSR contract, were designed to provide data on the relative effectiveness of prompting trials and anticipation (confirmation) trials, and on the effectiveness of partial cueing. The two experiments used different paired-associate materials and somewhat different experimental procedures and designs. The S-R pairs used in Experiment I were English words and one-digit numbers. In Experiment II, French words and their English equivalents served as the stimuli and responses.

The same two independent variables were investigated in each experiment. The first variable was number of initial prompting trials: in the four treatments under this variable, the subject had zero, 1, 2, or 3 prompting trials prior to responding by anticipation. The other variable was degree of correction provided following errors on anticipation trials. The correction procedures used ranged from a simple right-wrong indication to straightforward giving of the correct answer, with intermediate partial-correction stages in which S was given two or more alternatives and told that one of these was correct. Following partial correction, the subject either went immediately to the next item (Experiment I) or kept trying until the correct one of the alternatives had been chosen (Experiment II). Factorial designs were used in both experiments, so that all combinations of the several chosen values of each of the variables could be examined.

Results of Experiment I showed a highly significant effect due to prompting trials, a significant effect due to correction treatment in the case where no prompting trials had been given, and a significant interaction which reflects two clear facts: (1) A prompting trial was more effective than an anticipation trial when partial correction was employed but not when full correction was used, and (2) full correction was superior to partial correction when there had been no initial prompting trials but not when there had been two or more prompting trials.

In Experiment II, the advantage of prompting trials over anticipation trials (found in Experiment I) did not materialize, and there was in fact some indication of a trend in the opposite direction. The difference in the results of the two experiments was interpreted as being dependent on the nature of the correction treatment: In Experiment I, subjects did not respond overtly following partial correction; in Experiment II, they continued to choose from among the partial-correction alternatives until they made a correct response.

The several experimental treatments are discussed by the authors as training strategies, in terms of their absolute effectiveness in teaching the material, and in terms of their efficiency, considering both level of learning attained and the training time required. The conclusion is drawn that the prompting method can under suitable conditions be highly efficient for teaching tasks of the sort used in these experiments.

Angell, D., & Lumsdaine A. A. A study of subject-controlled partial cueing in paired-associate learning. San Mateo, Calif.: American Institute for Research, September 1961(c). Report AFOSR-TN-61-1342.

This study was performed, under contract with the Air Force Office of Scientific Research, to determine what effect a training procedure which utilized partial cueing at the option of the learner would have upon the learning of paired-associate materials. The technique was a simple one which provided partial cueing by successive revelation, upon student demand, of the letters of the three-letter response term in an S-R pair. No instrumentation was employed; the experimenter served as the "device" by which response components were revealed (aurally) to the student.

The study has potential applicability to auto-instructional programming in that the technique of partial cueing, if found effective, could be easily implemented in virtually any form of teaching machine, or in programmed texts. It is a cueing technique which, when applied to programmed material, is virtually content-free, not depending upon specific semantic or grammatical cues. However, the data showed little difference in overall effectiveness between the partial cueing technique and a standard anticipation procedure for learning paired associates. These findings are discussed in relation to other studies of partial cueing in paired-associate learning.

Fairbanks, G., Guttman, N., & Miron, M. S. The effects of time compression upon the auditory comprehension of spoken messages. Lackland Air Force Base, Texas: Air Force Personnel and Training Research Center, AFPTRC-TN-57-73 (ASTIA Doc. No. 131425), June 1957.

This report comprises a series of three studies conducted at the University of Illinois under Air Force Contract No. AF 18(600)-1059. The AFPTRC report deals with three experiments by Fairbanks, Guttman, and Miron which they reported serially in three separate articles in the March, 1957, issue of the Journal of Speech and Hearing Disorders (Vol. 22, No. 1): "Effects of Time Compression Upon the Comprehension of Connected Speech" (pp. 10-19); "Auditory Comprehension of Repeated High-Speed Messages" (pp. 20-22); and "Auditory Comprehension in Relation to Listening Rate and Selective Verbal Redundancy" (pp. 23-32). These three studies are referred to here as Experiments I, II, and III, respectively.

The general purpose of this research was to investigate the effects of rate of presentation and of varied repetition upon the comprehension of orally presented instruction. Three specific questions were posed: (1) How much does comprehension of factual information depend on rate of message presentation? (2) Which is more effective for comprehension of factual information: (a) a single presentation of a message (lesson) at a moderate rate, or (b) two presentations of the same message at twice the initial rate (50 per cent compression)? (3) Can the time available for message transmission be more effectively utilized by speeding up the presentation and using the time thus saved for introducing selected message "redundancy" or reiteration?

Each of the three experiments was concerned with one of these specific questions. The basic experimental training materials consisted of two extended expositions of technical information on meteorology, recorded by an experienced speaker at 141 wpm (words per minute). Two 30-item multiple-choice tests were used to measure message comprehension. Rate of presentation of the two basic messages was varied by means of a device developed by the senior author and two of his associates, W. L. Everitt and R. P. Jaeger, for effecting automatic "time compression" of tape-recorded messages. This time-compression process involved electronically "slicing out" and discarding very short portions (a few milliseconds each) of the recorded message. The device then automatically rejoins the remaining portions so they are abutted without discontinuity.

Experiment I: Comprehension as a Function of Degree of Time-Compression. Independent groups of 36 technical school students each were assigned to five experimental conditions involving presentation of the message with 0, 30, 50, 60, and 70 per cent compression (141 to 470 wpm). Each of these main experimental groups was composed of four equal sub-groups representing different levels of aptitude as indicated by Technical Specialist stanines. The sequence of events was: giving of standardized instructions; presentation of a two-minute sample message to familiarize the students with the "sound" of the subsequent message; presentation of first message; administering of test covering first message; presentation of second message; administering of test covering second message. A sixth group (the "test-only" group) was given the comprehension tests without having heard the basic message material. Scores were adjusted for estimated pre-knowledge on the basis of scores of the test-only group.

For the moderately lengthy messages used, the scores obtained indicated that comprehension fell off very slowly as rate of presentation was increased from 141 wpm to 282 wpm, but that beyond 282 wpm comprehension dropped off very rapidly. Since the scores obtained tended to be only slightly less at 282 wpm (50 per cent compression) that at 141 wpm (zero compression), it could be concluded that doubling the initial rate of presentation almost doubled message efficiency (number of items of information learned per unit of message time).

Experiment II: Auditory Comprehension of Repeated High-Speed Messages. The principal data for this experiment consisted of scores of the zero-compression group in Experiment I, together with scores of an additional group which was given a double presentation of the same message material compressed 50 per cent. For the double-presentation group, the experimental session consisted of standardized instructions, sample message, first message, first message, test on first message, second message, second message, test on second message. In each group there were 36 technical-school students having the same aptitude distributions.

As expected, giving two presentations resulted in a substantially greater amount of learning than just one presentation, both for messages at a relatively low rate (141 wpm) and messages at a relatively high rate (282 wpm). On the other hand, the utilization of the original time to present two 50 per cent-compression versions did not significantly increase overall comprehension scores over those for the condition in which the same amount of time was used for a single presentation of the original message. However, there was some indication that a gain achieved by compression-repetition for the men in the lower stanine levels may have been offset by a loss for those in the higher stanines.

Experiment III: Effects of Selected Verbal Redundancy. The basic message materials were increased in length by 43 per cent by splicing into the original blocks of content which were non-repetitive restatements of selected ("experimental") portions of the original message. This lengthened message was then compressed in time by 30 per cent, so that the presentation-time was the same as for the uncompressed original. Independent groups of 38 technical-school students (all at the seventh stanine level) were randomly assigned to four conditions: two versions (long and short) by two compression levels (zero and 30 per cent). For each of the four groups, the procedure paralleled that in Experiment I.

The augmented (or redundant) treatment of selected items of information was found to result in a substantial increase in comprehension scores for the augmented items, both for the compressed and uncompressed versions. However, this improvement on the augmented items was accompanied by an approximately equal decrease in comprehension of the non-augmented items. Thus, when the basic presentation was speeded up from 141 to 200 wpm and the time saved thereby was used for selective redundancy, the net result of the two counterbalancing effects for augmented and non-augmented items cancelled out. Thus, no significant improvement occurred in total comprehension scores, although, on the experimentally augmented items considered alone, higher scores were obtained for the augmented-compressed version than for the non-augmented, non-compressed version (which required the same basic total-presentation time).

Faison, E. W. J., Rose, N., & Podell, J. E. The effects of rest pauses during training-film instruction. Air Force Personnel and Training Research Center, Training Aids Research Laboratory, Unpublished Laboratory Note (mimeo.), TARL-LN-55-44, 1955(a).

Faison, E. W. J., Rose, N., and Podell, J. E. A technique for measuring observable audience reactions to training films. Air Force Personnel and Training Research Center, Training Aids Research Laboratory, Unpublished Laboratory Note (mimeo.), TARL-LN-55-45, 1955(b).

The research described by these two papers was concerned with evaluating the effect of rest-pauses interpolated at intervals throughout an instructional-film session. The rationale for the study was that, particularly under circumstances not conducive to attentiveness (fatigue, unfavorable temperature conditions, etc.), progressive loss in attentiveness and consequent reduced learning may be offset by introducing periodic short "breaks" into an even moderately lengthy instructional presentation. Such a procedure could operate to parallel the finding of generalized increments in learning of material not covered by specific practice, during intervals between sections of a film, obtained by Maccoby, Michael, and Levine in the second study reported in Chapter 19. A twenty-minute film on Ohm's Law, previously constructed for a study by Miller and Levine (1952) was used in each of two experiments. In the first experiment the film was shown to 38 classes of Air Force basic trainees; 19 of these classes were assigned to the experimental group (which had three 30-second rest breaks at intervals of about five minutes throughout the film) and 19 to the control group (no rest-pauses during the film). The criterion test was a 32-item multiple-choice test over the film's content.

The results on the post-film test in this experiment showed a significant difference ($P < .01$, one tail with 36 df) favoring the experimental group (mean: 59.2 per cent correct answers) over the control group (mean: 52.6 per cent). Differences between the experimental and control groups were significant for both the high-intelligence S's ($P \leq .05$, one tail) and low-intelligence S's ($P \leq .05$, one tail), both in favor of the experimental (rest-pauses) groups. Examination of the data indicated that the experimental-group S's tended to do better on nearly every item and not just on the items judged to be those most directly covered in the content presented immediately after the rest-pauses. This would of course greatly weaken the interpretation of the results in terms of increased attentiveness produced by the rest-break procedure, to the extent that the content items were independent. With a logically interdependent topic such as the one used, however, the correlation of test-item content with point-of-presentation in the film is in itself not clear cut, and this might explain the lack of clear relationship between test-item results and location of rest-breaks. The differences between the experimental and control groups were marginally significant ($P \leq .06$, one tail) for the morning sessions, and highly significant ($P < .01$, one tail) for the afternoon sessions (when increased fatigue and temperature would be expected to maximize the usefulness of the rest-breaks).

In a second experiment, 24 groups of 15 trainees per group saw the same film in a special classroom equipped for taking infra-red motion pictures of the audiences while they were viewing the film. Twelve groups had the rest-break procedure, while the other twelve saw the film straight through. Post-film test data showed essentially the same patterns as the first experiment, with mean scores on the 32-item test of 55.5 per cent correct for the rest-break groups and 50.2 per cent for the continuous-showing groups (t for the difference of 5.3 per cent was 3.95 with 22 df). Analysis of the audience-observation data yielded by the infra-red motion-picture records taken throughout the showings are of interest as an instance of an attempt to get at intervening response-patterns presumed to mediate the relationship between procedural or content variations and the terminal indices of learning obtainable from post-instruction tests. It had originally been intended that the infra-red cinematography technique would also be employed in subsequent studies to get at more complex aspects of mediating behavior during instruction; however, further use of the technique was abandoned, partly because of programmatic changes and partly because of the expense of the technique and the difficulty of obtaining reliable judgments of relevant student behavior. Various categories of behavior thought to be indicative of attentiveness were tried; in the present instance, the measure finally adopted was a simple count, taken at successive time intervals, of the number of students in the audience who were judged to be looking at the screen. (Other, more subtle indices generally failed to yield reliable scores when judged independently by several observers.) The attentiveness indices were obtained for the experimental and control groups at various points during the film-lesson showings by chopping up each infra-red film record of the audience into six-second segments and having these judged in scrambled order, without knowledge by the judges of which experimental treatment or which temporal segment of the instruction they were judging. Attentiveness indices based on the average percentage of time each subject was judged to be looking at the screen were then computed. The average percentages overall were 77.4 per cent for the rest-break group and 69.4 per cent for the continuous-showing group.

Data showing the comparative levels for an index of attentiveness at progressive points during the film have been previously presented by Faison in a section of an Air Force Technical Memorandum edited by Smith, Aukes, and Lumsdaine (1954). The index reported is an arbitrary one that is proportional to the amount of time students were judged to be attending to the film during each interval of the film. Approximate values of an index for successive intervals of about two minutes each during the instructional presentation are shown in the following tabulation, with the location of the rest breaks indicated by the vertical lines. (The numbers in the first row of the tabulation, labeled "Time," indicate the approximate duration of the film up to the end of each interval for which attentiveness data are reported.)

Time	5	7	9	11	13	15	17	20
Attentiveness Index:								
Rest-Break Group	61 |	59	53	43 |	50	44	35 |	55
Continuous-Showing Group	59	43	35	28	30	22	27	28
Difference	2	16	18	15	20	22	8	27

As seen in these data, the generally higher level of attention for the rest-break group at each time point after the introduction of the rest breaks appeared to be especially marked following the second and third breaks.

Gropper, G. L., Lumsdaine, A. A., & Shipman, Virginia. Improvement of televised instruction based on student responses to achievement tests. (Studies in televised instruction: Report No. 1) Pittsburgh, Pa.: American Institute for Research, March 1961.

This study was conducted under Grant No. 7-36-047 from the Educational Media Branch of the U. S. Office of Education under Title VII of the National Defense Education Act. The instructional effectiveness of a preview version of a lesson was compared with that of a subsequent version which had been revised on the basis of student responses to each item of an achievement test, administered to students who had seen the preview version. Analyses of test items revealed which points were not effectively taught, and lessons were then revised accordingly.

At a later date, preview and revised versions of the lesson were simultaneously telecast to new groups of students over two different channels in Pittsburgh. Students who watched each version took identical tests. The experiment was replicated using different lesson content (a lesson on "heat" and a lesson on "chemistry"), different instructors, and different classes of students. Results showed improvement on various points of content ranging from six to twenty-six per cent for students who watched the revised versions over students who watched the original, preview versions. Most of the experimental comparisons made were statistically significant. The main importance of this study was in furnishing a concrete demonstration of how student responses to a test can enable a television instructor to tailor his presentation to the needs and abilities of his audience, much as the responses obtained in trying out a self-instructional program can be used as a basis for revising and improving it. A limitation of the study was that only one stage of revision was used for each lesson, as compared to the several stages of successive revision sometimes employed with self-instructional programs. A further limitation, inherent in situations where a pre-determined set of content must be presented in a lesson of fixed, pre-determined length is that results of test data from preliminary versions may indicate need for more instruction than is possible within the pre-allocated amount of time.

Guthrie, P. M., & Lumsdaine, A. A. Some effects of graduated partial cueing on the learning of paired associates. San Mateo, Calif.: American Institute for Research (AFOSR-TN-61-1341), December 1961.

This report describes a series of small-scale exploratory experiments conducted under an AFOSR contract to investigate the effects of several cueing methods on paired-associates learning. In these experiments, subjects learned the items in several lists of city names paired with their corresponding three-letter airline codes, using special cueing procedures or standard anticipation procedures. The cues consisted in presenting the response terms under lowered conditions of visibility. Visibility, and hence cue strength, was varied tachistoscopically or gradually by adjusting illumination of the cue. Under conditions in which subjects had increasing amounts of time to anticipate the response terms, tachistoscopic cueing produced reliably better learning than a comparable anticipation condition. However, the positive effects produced by this form of cueing were, though statistically significant, relatively small; furthermore, it was not found to be superior to the use of immediate full prompting for each term. Several other forms of cueing investigated yielded non-significant differences. The work represented by these exploratory experiments is primarily of interest in relation to the design of further experimentation on similar cueing techniques for use with paired-associate and continuous-discourse materials.

Kanner, J. H., & McClure, A. H. Varied versus identical repetition in filmed instruction on micrometer reading. Air Force Personnel and Training Research Center, Training Aids Research Laboratory, Unpublished Laboratory Note (mimeo.), TARL-LN-56-11, April 1956.

This study comprised two related experiments which were designed to compare the effects of "identical" repetition (in which identical examples are presented repeatedly) and of varied repetition (in which repetition utilizes similar but not identical examples) in the learning of a skill in which terminal performance comprises additional examples not identical to those practiced during the training procedure. The task to be learned was micrometer reading. Two sets of illustrative micrometer settings were used in the filmed instruction: (A) .175, .163, .498, and .773; and (B) .725, .339, .649, and .198. In each experiment, three conditions were compared: (AB) varied repetition (with the film containing eight different examples, or two different sets ["A" and "B"] of four examples each); (AA) identical repetition (the four "A" examples being repeated after each had been shown once); and (BB) identical repetition (the other ["B"] set of four examples repeated twice). The posttest consisted of 24 multiple-choice items, each presenting a photographed micrometer setting with four possible readings on a slide exposed for 20 seconds. However, only sixteen of these items were scored; these were 16 micrometer settings not used in the training film for the varied-repetition group.

In Experiment I, each condition was represented by two groups, "advanced" and "regular" (as defined by a mathematics proficiency test score); these two kinds of groups received their instruction separately, in accordance with normal procedures then in force at the training center. The procedure for the advanced groups differed slightly from that for the regular groups; the latter were told in advance that they would be tested subsequently, while the former (advanced groups) were not told. In addition, the advanced groups were tested one week following the learning session, while the regular groups were tested immediately after the film. These procedures were used to reduce the criterion test scores of the advanced group *S*'s in order to avoid a ceiling effect which could act to mask any differences in the effects of the two experimental treatments. (Pre-film testing had indicated that the advanced *S*'s would score very high.) Subjects were Air Force basic trainees from 36 flights (of about 50 men each) and were assigned in a pre-determined random order to treatments AB (18 flights) and treatments AA and BB (nine flights each). The results showed no significant differences between the identical- and the varied-repetition conditions for the regular groups; for the advanced groups, however, there was a significant difference in favor of the varied condition over the average for identical-repetition conditions ($t = 2.92$, 11 df, $p = .02$). The means are shown in Table 1.

A tentative conclusion from the first experiment was that the potential advantage of using a more varied set of examples might operate only for the more intelligent trainees, whereas added practice with the same examples might be as helpful or more helpful for the less intelligent.

TABLE 1

Mean Number of Micrometer Settings Given Correctly for "Varied" and "Identical" Treatments by "Advanced" and "Regular" Groups in Experiment I (16-item multiple-choice post-film test)

	Identical (AA)	Identical (BB)	Mean AA & BB	Varied (AB)	Diff. (V - I)
"Regular" Group (Immediate Testing)	10.2	10.4	10.3	9.9	-0.4
"Advanced" Group (Delayed Testing: 1 week after film)	11.2	12.3	11.8	12.8	1.0

Experiment II was designed as a replication to investigate further this supposition. The experimental procedure was similar to that of Experiment I, and used the same training film and test. The Air Force basic trainees from 48 flights were randomly assigned by flights to treatments, 24 flights to the "AB" (varied-repetition) film and 12 flights each to the "AA" and "BB" (identical-repetition) treatments. Half of the flights were tested immediately after the instructional film (and had been told before the film that they would be so tested). The other half of the flights were not told they would be tested and were given their test after a two-day delay (instead of the one-week delay used in the first experiment). Because of changes in normal procedure at the training center during the interval between the two experiments, there was no separation into "advanced" and "regular" groups. Instead, analysis in terms of intelligence was accomplished by splitting each flight into "low" and "high" sub-groups on the basis of AFQT scores after the experiment was completed. The mean scores on the 16-item criterion test are shown in Table 2. The results indicated no significant difference among the conditions both for the high- and low-intelligence groups.

Differences between Identical and Varied treatments in Experiment II were non-significant in all cases, for both AFQT sub-groups under both immediate- and delayed-testing conditions. The second experiment thus failed to provide verification for the expectation that the relative merits of the two treatments would operate differentially for the sub-groups of differing intelligence.

TABLE 2

Mean Number of Micrometer Test Settings Given Correctly for "Varied" and "Identical" Treatments by High- and Low-Intelligence Sub-groups in Experiment II

		Identical (AA)	Identical (BB)	Mean AA & BB	Varied Film	Diff. (V - I)
Low-Intelligence Groups	Immediate Testing	10.54	11.21	10.88	10.67	-0.21
	Delayed Testing*	9.52	9.76	9.64	10.06	0.42
High-Intelligence Groups	Immediate Testing	14.31	14.35	14.33	14.46	0.13
	Delayed Testing*	12.77	13.25	13.01	13.29	0.28

* Two days after film.

Kimble, G. A. The effect of praise, reproof, and no incentive on the value of a training film. Air Force Personnel and Training Research Center, Training Aids Research Laboratory, Unpublished Laboratory Note (mimeo.), TARL-LN-55-7, December 1955(a).

This study and the following one (Kimble, 1955b) were conducted as adjuncts to the experiment reported in Chapter 15. It was designed to compare the effects of "praise," "reproof," and "no-incentive" conditions upon learning. A 2 x 2 x 3 experimental design was used, varying (1) ability (high vs. low AFQT scores), (2) participation conditions (guided versus unguided--cf. Chapter 15), and (3) incentive conditions (praise, reproof, or no incentive). An earlier study on incentive conditions (Hurlock, 1925) was considered by the author to be inconclusive due to confounding factors (the "praise" group appears to have received greater urging to do better than the other groups, in addition to being praised).

The learning session consisted of two films. The first dealt with how to read a slide rule; the second dealt primarily with how to use the slide rule to multiply and divide (as well as how to read it). A test on reading of the slide rule (25 items) was interpolated between the two films, followed immediately by the incentive treatments, which (for the two groups to which they were given) consisted of a brief statement given by a specially trained non-commissioned officer after observing the trainees performance as they were taking the slide-rule reading test. For the praise condition, he told the trainees that most of them seemed to be doing well and at least up to standard on a "subject which is pretty difficult" and that their performance was "very gratifying." For the reproof condition, he gave exactly the same message with such substitutions as doing poorly, "far below standard," and "very disappointing." In neither condition was anything said about trying to do better or paying closer attention. The second (final) test, which followed the second film, consisted of the same 25 items on how to read the slide rule, plus 10 items on how to use the slide rule. Subjects were Air Force Basic trainees, who were assigned to treatments by quarter-flight groups as explained in Chapter 15. Each film-showing group, or audience, was composed of four independent quarter-flights who were instructed and tested as a single group. Within each of the treatment-cells defined by incentive and participation condition there were thus two showing groups which could be considered as replications. Selective discarding was used to match up the AFQT distribution for one of these, across incentive conditions, as a relatively "high" AFQT group (mean AFQT, 65.7) and to match up the other as a relatively "low" AFQT group (mean AFQT, 55.3). For each incentive condition this yielded 383 S's comprising four independent replications, paired on AFQT status and participation condition; using these as units of sampling and analysis gave a highly conservative test with 3 df for comparing any two incentive conditions.

On the items which tested how to read the slide rule, no significant differences were found among the incentive conditions in the gains from first test to second test (see Table 1). However, on the 10-item test on how to use the slide rule, there were significant differences among the incentive conditions, with reproof being highest, and no incentive being lowest (see Table 2).

TABLE 1

Mean Gain Scores (Second Test Minus First Test)
For Reading of Slide-Rule Scales

AFQT	Participation Condition	Incentive Condition: Praise	Reproof	None
High	A	2.17	1.67	1.80
	B	2.97	1.80	3.10
Low	A	2.09	2.24	1.88
	B	1.44	3.41	2.67
Mean		2.17	2.28	2.36

TABLE 2

Mean Scores on Slide Multiplication and Division Test

AFQT	Participation Method (cf. Ch. 15)	Incentive Condition: Praise	Reproof	None
High	A	3.70	3.70	3.50
	B	3.73	4.47	4.03
Low	A	2.74	3.26	2.41
	B	2.79	3.15	2.41
Mean		3.24	3.64	3.08

None of the lower-order interactions were significant when tested against triple interaction; using pooled interaction (7 df) the estimate of error, the value of F for incentive conditions was 5.50 (P: .05); the values for t and P (one-tail, 3 df) among pairs of treatments were t = 2.29 (P: .05) for reproof versus praise, t = 3.14 (P: .03) for reproof versus no incentive, and t = .85 (P: .25) for praise versus no incentive. Since the training on how to use the slide rule followed (while the training on how to read the slide rule preceded) the incentive treatment, the data show no evidence that the incentive communications lead to any improvement in a skill which had already been partially mastered (reading the slide rule), but did have an effect on the mastery of a new skill (using the slide rule).

Kimble, G. A. The value of an interpolated test in increasing the effectiveness of a training film. Air Force Personnel and Training Research Center, Training Aids Research Laboratory, Unpublished Laboratory Note (mimeo.), TARL-LN-55-10, December 1955(b).

Two experimental comparisons were made to investigate the effect of interpolating a test (without knowledge of results) between two portions of an instructional session. (These experiments, like the preceding one reported by Kimble [1955a], were conducted as adjuncts to the experiment reported by Kimble and Wulff in Chapter 15.) One purpose of the experiments was to investigate the extent to which such a test may act (1) to increase motivation and general attention (cf. Chapters 18 and 19) and (2) to show the student what the instructor regards as important or to indicate to him what he knows and what he does not know. Two training films were used. The first one was a slide-film (see Chapter 15) which dealt only with how to read a slide rule; the other (cf. Kimble, 1955a) was a U.S.O.E. film which dealt also with the use of the slide rule for multiplication and division. The 20-minute interpolated test consisted of 25 items testing how to read the slide rule (see Chapter 15); the final test (all S's) included 10 items on the slide rule's use (as well as 25 new items on how to read it). The interpolated test, when given, immediately followed the first film and immediately preceded the second film. No compensating break was given the men in the no-test condition; after the first film they proceeded immediately to the second film. For the multiplication and division items, subjects were given slide rules and were required to do two things: (1) to set up the slide rule to solve a problem and then to match the position of the slide rule to one of six drawings on the test form (each of which showed a different possible position of the slide); and (2) to write down the answer to the problem as read from the slide rule. In scoring the test, one point of credit was allowed for the correct setting. If the setting was correct, one additional point of credit was given if the correct numerical answer had been recorded. No credit was allowed for a correct answer if the setting was incorrect, because this implied that the problem had been worked out on paper without the use of the slide rule.

Experiment I employed basically a 2 x 2 design, the variables being: (1) interpolated test versus no interpolated test, and (2) guided versus unguided participation (cf. Chapter 15). The subjects were approximately 500 Air Force basic trainees comprising 15 men randomly chosen from each of 32 flights of 60 men each. Eight of these quarter-flights were assigned to each of the four basic treatment cells; thus there were eight replications in each cell, each consisting of one independently sampled quarter-flight. For analysis, the eight quarter-flights within each treatment (cell) were arranged by the rank order of their mean AFQT scores, so as to comprise eight AFQT-stratified quartets; the units of analysis used were thus quarter-flight means in a 2 x 2 x 8 design. (Since the instruction and testing were done in showing-groups of four quarter-flights each, any variations in showing-group procedure could have affected groups of four quarter-flights as a whole; an even more conservative approach statistically might therefore have considered an entire showing-group of four quarter-flights as one unit for analysis, with two replications to each of the four cells, and total df = 7—cf. Kimble [1955a], and Chapter 15.)

Table 1 presents the means on the final slide-rule scale reading test for the quarter-flights, arranged in ascending order of their mean AFQT scores, for the interpolated-test and no-interpolated-test groups who were instructed under guided (A) and unguided (B) conditions. Analysis of variance using a residual error term with 22 df yielded an F of 4.55 (P: .05) to indicate a significant effect of the interpolated-test procedure. The F of 4.44 for participation condition was also significant at the .05 level on this second test: cf. results on first test, given in Chapter 15. The variation due to quarter-flight differences in AFQT was not significant, however (F = 1.84), suggesting that stratification on the basis of AFQT means may have added little to the sensitivity of the comparisons. The corresponding data for the final test on use of the slide rule for multiplication and division are presented in Table 2.

TABLE 1

Mean Slide-Rule Reading Scores for Quarter-Flights on the Final (Second) Test (Experiment I)

Mean-AFQT Rank:	1	2	3	4	5	6	7	8
Interpolated Test								
(A)	16.21	18.79	17.47	19.47	19.48	19.88	15.67	18.93
(B)	14.60	18.45	18.33	19.07	14.53	17.55	17.75	17.50
No Interpolated Test								
(A)	17.17	15.82	14.93	15.43	18.40	19.46	19.00	17.44
(B)	13.65	15.55	13.12	17.64	16.14	17.23	17.08	17.31

TABLE 2

Mean Multiplication and Division Scores for Each Quarter-Flight on Final (Second) Test (Experiment I)

Mean-AFQT Rank:	1	2	3	4	5	6	7	8
Interpolated Test								
(A)	1.21	3.50	2.60	4.07	2.93	2.58	4.62	3.71
(B)	1.40	5.18	3.27	3.00	2.20	2.00	2.88	3.25
No Interpolated Test								
(A)	1.33	2.12	4.80	1.36	5.60	3.77	6.00	3.61
(B)	2.12	0.82	0.94	2.36	1.57	0.33	2.54	4.15

A parallel analysis of variance for the data in Table 2, on how to <u>use</u> the slide rule, failed to yield a significant $\underline{F}$ value for the interpolated-test variable ($\underline{F}$ = 0.45). The results might suggest that an interpolated test has a specific effect rather than a general one—i.e., it may improve performance only on material which it specifically covers. It should be noted that the data on slide-rule use (not covered by the interpolated test) may lack sensitivity due to the low level of performance attained (mean scores were 3.02 for the interpolated-test group and 2.71 for the other group, out of a possible 20 points). There was, however, a significant difference between the participation conditions ($\underline{F}$ = 4.53, $P < .05$), but the author suggests this may have been due to a confounded effect from unequal AFQT scores.

Experiment II employed a two-group design, comparing the effects of interpolated versus no interpolated test. Again, each group was divided into eight strata on the basis of AFQT scores. Subjects were approximately 200 Air Force basic trainees, comprising 16 quarter-flights which were used as the units of sampling and analysis. In this experiment, one film—the U.S.O.E. film dealing with how to read and use a slide rule—was used twice, once before and once following the interpolated test (which, in this case, dealt with both the reading of the scales and their use). The no-interpolated-test group saw the same film twice without a pause. On the same final test as the one used in Experiment I, no significant differences were obtained on how to read the slide rule (the means being 13.82 for the interpolated-test group, and 13.87 for the no-interpolated-test group). On the part of the test dealing with how to use the slide rule, the corresponding means were 3.51 and 3.19, and the F ratio between the two groups was marginally significant ($F = 3.62$, $P = .08$, two tailed) for an analysis of covariance with test scores adjusted on the basis of AFQT scores.

Miller, J., and Klier, S. The effect of interpolated quizzes on learning audio-visual material. Air Force Personnel Training and Research Center, Training Aids Research Laboratory, Unpublished Lab. Note (mimeo.), TARL-LN-56-8, January 1956(a).

This study and the following one were conducted at Queens College under Contract No. AF 18 (600)-38. This study was primarily concerned with the effects of interpolated activity between successive showings of a training film and also compared single versus multiple presentations of the film. The rationale was that the increments due to repetition of the film could be increased by interpolating a quiz between each pair of showings, and that the gain should exceed that due merely to the concomitant increase in distribution of practice. Eight different flights of Air Force basic trainees were tested after each of the following five treatments: (1) "spaced quiz"—quizzes interpolated between each of the three showings of the same film ("Principles of Lubrication"); (2) "spaced"—an unrelated task interpolated between the three showings of the same film, in order to space the three presentations; (3) "massed"—the three showings of the same film with no intervals between them; (4) one showing of the film; and (5) no film. Subjects were assigned (also instructed and tested) in flight units, and thus a flight was properly considered to be the sampling unit in the statistical analysis. The criterion test consisted of 26 multiple-choice items on the film content, of which the first nine were identical to those used in the quizzes for the spaced-quiz group. Results on the first nine items showed that the spaced-quiz group was not significantly superior to the spaced group which had had no prior exposure to these items (although superior to the other groups); thus, practicing the items (without knowledge of results) did show reliable evidence of better learning. Results were also analyzed both for the total test (26 items) and for the 17 items not included in the quizzes. Data summarized here are for the total test.

Comparison of one-showing mean score (12.9) and the no-film mean (7.9) showed a highly reliable though not numerically large gain from seeing the film (t = 7.5, with 14 df). In the analysis, some attempt was made to reduce variability by matching up flights of similar distribution on ability-test (AFQT) scores; however, full use of the stratifying (AFQT) data was not made, and the complete data were no longer available when this summary was prepared. This might account in part for the lack of significant inter-treatment differences, using 14 df in each instance. The superiority of the condition with the highest obtained mean (spaced quiz; mean, 14.8) was of borderline significance (t = 2.5) as compared with the one-showing group (mean, 12.9), while t in comparing the latter with the mean for the three-showing massed group (13.4) was only 0.6. The obtained mean of 14.5 for the spaced (no-quiz) group was nearly as high as for the spaced-quiz group; none of the three-showing groups differed significantly from the others. Thus, though the ordering of the obtained means was consistent with expectation, the analysis was not successful in demonstrating the advantages postulated for the interpolated quizzes. (Cf. Kimble, 1955b; and Romney, 1952).

Miller, J. and Klier, S. The effect on active-rehearsal types of review of massed- and spaced-review techniques. Air Force Personnel Training and Research Center, Training Aids Research Laboratory, Unpublished Lab. Note TARL-LN-56-6 (mimeo.), January 1956(b).

This study was conducted as a follow-up to the studies by Miller and Levine (1952) and by Miller, Levine, and Sternberg (1956). These two studies, both using high school students as subjects but employing two different film topics, found "massed review" superior to "spaced review." (A replication of the first study with the same film shown military subjects was, however, unsuccessful in replicating the Miller and Levine results: see summary of report by Miller, Levine, and Sternberg, 1954). The present study was concerned with whether the superiority of massed over spaced review would apply to active review (in which students respond overtly to review questions instead of merely having the review presented to them) as well as to the "passive" types of review previously employed. It also made a direct comparison between the active- and passive-review procedures, using the same test and same film—"Pattern for Survival"—employed in the 1956 Miller, Levine, and Sternberg study (and also, earlier, by Michael and Maccoby: see Chapter 18). Subjects were high school students from 34 classes (sampled by class groups). One class from each of eight different high schools was assigned to each of three review conditions: (1) active massed, (2) active spaced, and (3) passive massed; and one class apiece from six of the schools to a control group which was tested without seeing the film.

Comparisons of immediate post-test scores showed highly significant differences (using t-tests) between each of the experimental groups (means of about 138) and the control group (mean of 40), indicating that substantial learning occurred from the film-review combinations; but no significant differences were found among the three experimental review groups on immediate testing (t was less than .05 in each case). When the classes that were still available were re-tested for retention after a three-week period the means were 123 for active spaced, 117 for passive massed and 117 for active massed; the largest t, for the first two, was 1.4 (10 df). It may be noted that lack of an advantage for massed over spaced review with the active form of review could arise from the fact that, despite the use of KCR, potential benefits of spaced practice in the massed-review condition may be overbalanced by the effect of the students practicing more errors in the massed-review condition (which is less well prompted because temporally more remote from the corresponding first presentation of each topic).

Miller, J., and Levine, S. A study of the effects of different types of review and of "structuring" sub-titles on the amount learned from a training film. Washington, D. C.: U. S. Air Force Human Resources Research Laboratories, HRRL Memo. Report No. 17, March 1952.

This is the first of several reports dealing with the effects of review conditions on learning from a training film. These studies were conducted at Queens College, New York. Herman Fish and Joseph Kanner assisted in this study, conducted under Air Force Contract No. AF 33 (038)-13507. It was concerned with ascertaining the effects of (1) variation in review conditions and (2) sub-titles which serve to structure the contents of the film for the student. Several versions of an experimental film on Ohm's Law were constructed for the experiment. The film was approximately 20 minutes in duration and contained four sections, dealing successively with the topics of voltage, resistance, current, and the relation among these as given by Ohm's Law. In addition to the main presentation, averaging around four minutes per topic, review sections of a minute or so per topic were included. The presentation sections were identical in all film versions; review sections were identical except as noted below. The basic variation was in the placement of the reviews. Under "spaced-review" conditions, each review immediately followed the corresponding presentation section, so that each review was separated from review of other topics by an intervening presentation section. In the "massed-review" condition, by contrast, all presentation sections were first given prior to any review, and the four review sections were massed together at the end of the film.

A 4 x 3 experimental design was used. There were four review conditions: (a) massed review, (b) spaced review, (c) spaced oral review (only the narration on the film sound track was used without the visual material), and (d) no review. Three conditions of sub-titling were employed: (a) major sub-titles only (these sub-titles identified and demarcated the four major topics of the film); (b) complete sub-titles (minor sub-titles were used, in addition to the major ones, to demarcate approximately four sub-topics within each of the four major topics of the film; and (c) no sub-titles. The subjects were 36 classes of high school students (approximately 30 students per class). Three classes were assigned as units to each of the twelve cells; each class was then instructed and tested as a group. The class rather than the individual was therefore taken as the unit of sampling and analysis. All classes were first given a pre-film test, then were shown the film and given a post-test. Two parallel test forms were used, each consisting of 15 multiple-choice items. Each of the two test forms was used as pre-test for half the classes and as post-test for the other half. Although randomization was relied on as a basic procedural control, some increase in sensitivity was sought by stratification in terms of IQ scores; extreme cases were discarded to bring the mean IQ for each class to a value of 106. Overall mean scores were about 3.4 for pre-test and 7.7 for post-test, with the average increment in knowledge represented by the gain of 4.3. Percentages correct on individual test items ranged from 16 to 76 per cent, with about half the items in the 50-60 per cent range. Pre-test scores evidently did not exceed expected chance level; post-test scores only were therefore used for most of the analysis of results.

Analysis of variance showed an F of less than unity for sub-title conditions and an F of 2.66 for review conditions (3, 24 df). In computing t-tests among review conditions, sub-title variations were ignored. The significant comparisons are shown in Table 1. Only the massed-review condition was shown to be significantly superior to the no-review condition.

TABLE 1

Comparison of Mean Post-Scores for the various Review Conditions
(Curtailed-range Sample)

Groups Compared	M	df	t	P (both tails)
No Review Massed Review	7.36 8.87	16	2.65	.01 to .02
Spaced Review Massed Review	7.49 8.87	16	2.49	.02 to .05
Spaced Review—Oral Only Massed Review	7.49 8.87	16	2.60	.01 to .02

Similar results were found when closer matching of groups was carried out. Analysis by separate film topics showed that the direction of advantage for massed review over spaced review was consistent for all four topics.

Two factors were considered relevant in accounting for the superiority of massed over spaced review. One factor is distribution of practice, i.e., massed review distributes practice on each topic and hence may be regarded as a spaced-practice condition (present A, B, C, D; then review A, B, C, D), while spaced review masses practice (present A, review A; present B, review B; etc.). Another factor may be recency. With the massed-review procedure, the criterion test follows the reviews immediately, while with the spaced-review procedure, there is a longer interval following most of the review portions of the film before the criterion test.

Miller, J., Levine, S., and Sternberg, J. The effects of different kinds of review and of sub-titling on learning from a training film: a replicative study. Air Force Personnel and Training Research Center, Training Aids Research Laboratory, Unpublished Staff Research Memorandum (mimeo.), June 1954.

This experiment was an unsuccessful attempt to replicate a previous study by Miller and Levine (1952), in which massed review was found to be superior to spaced review, and also to control the variable of recency. The massed-review procedure was followed immediately by a post-test in the Miller-Levine study, while a greater time interval elapsed between most of the spaced review and the post-test. To check on this, the present study introduced the additional variable of delay (by introducing an irrelevant film after the presentation of the experimental film). The experiment employed a 2 x 2 x 3 design, with two conditions, delay and no delay, for each combination of sub-titling (sub-titles versus no sub-titles) and review (spaced, massed, or no review). Subjects were 36 classes of Air Force basic trainees, instead of the high school students used previously. Three classes were assigned to each of the 12 cells. An additional control group of five classes was tested without being shown the film. The results showed significant effects of the film in comparing all experimental groups with the no-film control group, but no significant differences were found for the variable of sub-titling (as before), nor for review conditions (unlike the previous study). The largest t-value for any comparison among experimental group means was about 1.2. However, little import can be attached to these negative results, particularly since the class means were highly variable and even the differences between immediate test and delayed test were small and non-significant (as were those between either form of review and no review at all). Thus the conditions necessary for a satisfactory experimental test of the hypotheses did not obtain. The experimenters suggest that this situation may have been due, in turn, to (1) weak motivation (e.g., the trainees had been subjected to a battery of tests prior to the experiment) combined with excessive auditory and visual distractions which were present during testing, and (2) the rather uninteresting nature of the film which was used.

Miller, J., Levine S., and Sternberg, J. Extension to a new subject matter of the findings on the effects of different kinds of review on learning from a training film. Air Force Personnel and Training Research Center, Training Aids Research Laboratory, Unpublished Lab. Note TARL-LN-56-4 (mimeo.), January, 1956.

This experiment was, like the preceding one by the same authors, conducted at Queens College under Air Force Contract No. AF 33(600)-38. It was a second follow-up to the study by Miller and Levine (1952), and replicated the comparison of "spaced" versus "massed" review with a new subject matter. Since the failure of a first attempted replication—see Miller, Levine, and Sternberg (1954)—was attributed in part to the inability of the training film to hold the attention of the subjects, a film judged to have optimum interest was used ("Pattern for Survival"—dealing with defense against atomic attack). Massed-review, spaced-review, and no-review conditions were compared in this experiment, just as in the previous (Miller and Levine) study. Subjects were 23 classes of high school students (about 20 per class). Eight classes each were assigned the no-review and spaced-review conditions; seven classes were assigned to the massed-review condition. Subjects were instructed and tested in classes; the class was taken as the sampling unit. The results verified those reported earlier by Miller and Levine, showing that: (1) massed review was superior to spaced review, and (2) massed review was superior to no review; spaced review was also superior to no review in this study (see Table 1).

TABLE 1

Comparisons of Test Scores for the Various Review Conditions

Groups Compared	M	df	t	P
Massed Review	139.36	13	2.40	.05
Spaced Review	125.26			
No Review	101.02	13	6.72	.001
Massed Review	130.36			
Spaced Review	125.26	14	3.79	.01
No Review	101.02			

Romney, A. K. Effects of selected "set"-inducing instructions upon learning from multiple film showings. Boston, Massachusetts: Boston University, unpublished report, Contract No. AF 33(038)-22944, September 1952.

This experiment attempted to investigate the effect of set-inducing instructions on learning from repeated showings of a film. Because of the rather small N's and small size of differences obtained, the study must be regarded as an exploratory one. The experimenter tried to create two different "sets" to learn two contrasted kinds of content from either two or three showings of a film called "The Airplane Changes the World Map." "Principle" set-inducing instructions consisted of giving sample questions, before a film showing, which required an understanding of general concepts underlying the topics presented in the film; "fact" set-inducing instructions consisted of pre-film sample questions involving the rote recitation of some factual content of the film. Subjects were summer-session college students. Subjects in each of four classes were assigned randomly to one of six treatments employing two showings of the film; subjects in two other classes were randomly assigned to one of three additional treatments in which the film was shown three times. Immediately after their last showing, subjects were given an orally presented, short-answer, write-in test comprising 16 "fact" items and 15 "principle" items. All groups were told before seeing the film that they would be tested and should try to learn as much as possible from the film. Random assignment of subjects within a classroom group was made possible by giving the "set" instructions in booklets passed out randomly to the subjects. Most of the S's were males; because of sizable sex differences, the experimenter limited the analysis presented to the male S's.

The main data for the nine groups are summarized in Table 1, in which "n" stands for no "set" instructions, and "F" and "P" respectively stand for "Fact" and "Principle" set-instructions preceding each of the two or three film showings. (The "FP, FP" group was given both "set" instructions before each showing; the "F, P" S's got the fact set before the first showing and then the "P" set before the second, etc.)

TABLE 1

Summary of Experimental Conditions, with N's and Mean Test Scores for "Fact" Items, "Principle" Items, and Total (Males Only)

	Two-showing Groups						Three-showing Groups		
Group No:	1	2	3	4	8	9	5	6	7
"Set":	n,n	F,F	P,P	FP,FP	F,P	P,F	n,n,n	n,n,F	n,n,P
N (males)	17	20	12	12	16	14	11	13	13
Means:									
"Fact"	8.8	10.0	8.9	10.3	8.4	8.9	9.8	11.2	9.7
"Principle"	6.8	5.8	6.8	7.1	5.9	6.5	7.0	7.2	7.4
Total	15.6	15.8	15.7	17.4	14.3	15.4	16.8	18.4	17.1

The comparisons that can be reported are somewhat limited by the incomplete availability of data. The standard error of the difference between means was close to unity for each of the comparisons between groups for "Fact" or "Principle" scores separately, so that as a rough index t approximately equals the mean difference between any two groups. A rigorous comparison with the data available is possible only within two-showing or within three-showing conditions because of the sampling scheme used; however, this fact is of consequence in relation to only some of the relevant comparisons.

The first question of interest is whether such results as those of May and Lumsdaine (1958, Ch. 7) are replicated in terms of an increase in learning for the kind of content stressed by the pre-film-set material at the expense of lesser learning of other material. Comparison of obtained test means for Group 2 versus Group 1 (F, F versus n, n) shows this pattern as a result of the "Fact" set instructions. An advantage for Fact set in increasing Fact scores is suggested in comparing Group 4 versus Group 3 (FP, FP versus P, P) and Group 6 versus Group 5 (n, n, F versus n, n, n). The values of t for advantage on Fact scores are only 1.3, 1.3 and 1.2, respectively, for these comparisons; however, an overall significance level of .025 is given for the combined results. For possible effects of Principle set on Principle scores a similar comparison was made for Group 3 versus Group 1, 4 versus 2, and 7 versus 5; here, the only t greater than unity was that of 1.5 for Group 4 versus Group 2 (FP, FP versus F, F). An effect of the "Principle" set may, however, be seen when subjects are divided according to ability into high and low groups (on the basis of the Wonderlic Personnel Test). Combined results for Groups 3 and 4 versus Groups 1 and 2 then show an advantage on the Principles score of 2.05 (9.09 versus 7.04) for the higher-ability subjects as against a difference of only 0.08 (5.15 versus 5.07) for the lower-ability subjects.

A further question of interest was whether the total effects of the fixed presentation given by a film could be increased by concentrating students' attention on one aspect during a first showing and a different aspect during a second showing. That this attempt seems to have been unsuccessful with the procedures and materials here used is suggested by comparing the obtained total scores of Groups 8 (F, P) or 9 (P, F) with those of Group 1 (n, n) or with those of the other two-showing Groups (Groups 2, 3, 4).

A third question of interest in designing the study was related to previous findings (see also Chapter 17) showing diminishing returns from successive repetitions of film material; e.g., McTavish (1949) found no advantage for three showings over two. The supposition here was that gains might, however, be found if a new "set" were introduced to concentrate attention on a special aspect of the material after the second showing. Comparison of the mean Fact score of 11.2 for Group 6 (n, n, F) with those of 9.8 and 8.8, respectively, for Group 5 (n, n, n) and Group 1 (n, n) gives only limited support to this possibility, since t for Group 6 versus Group 5 is only 1.5, and the significance of the apparent superiority of 2.4 points for Group 6 over Group 1 cannot be determined precisely with the data available.

Weiss, W., and Fine, B. J. Stimulus familiarization as a factor in ideational learning. Air Force Personnel and Training Research Center, Training Aids Research Laboratory, Unpublished Lab. Note TARL-LN-55-48 (mimeo.), December 1955.

This study was conducted at Boston University under Contract No. AF 33(038)-22944 with the U. S. Air Force Human Resources Research Laboratories. It was concerned with investigating whether or not prior familiarization with key terms in a film will facilitate learning from the film. Three groups were used: (1) Familiarization group: subjects were given 11 minutes of paired-associate training on naming various subdivisions of the United Nations; immediately following this they were shown an 11-minute sound film-strip, "Structure for Peace," which dealt with the UN, but which they had not previously been told they would see. (2) Two-showing group: This control group was shown the film twice, successively; this procedure took as much time as the familiarization group, and thus afforded a control for the time variable. (3) One-showing group: this group was shown the film once. Immediately following the session, all groups were given a 20-item multiple-choice test on the structure and functions of the six main parts of the UN. Five additional questions on the UN covering general knowledge not presented in the film were also asked to give a rough index of pre-knowledge about the UN. The subjects were 700 high school students from 28 classes. Sampling was by classes; thus, in the statistical evaluation, the class was properly taken as the sampling unit. Results are presented in Table 1. It can be noted that: (1) There were no significant differences among the groups on pre-knowledge of the UN; (2) the familiarization group learned significantly more than the one-showing group ($P < .05$, one tail); and (3) the two-showing group learned more than the one-showing group ($P < .05$, two tails), as expected. However, no significant difference was found by the investigators between the familiarization group and the two-showing group. Prior familiarization thus appeared to be about as effective as a second exposure of the learning material. The students were also asked, after seeing the film, to rate its interest to them. The obtained interest ratings were lower for the familiarization group, though the test of significance used did not demonstrate this difference to be reliable.

TABLE 1

Mean Number of Fact-Quiz Items Correct

Group	N (Classes)	General Information about UN (5 items)		Content of Film-Strip (20 Items)	
		Mean	SD	Mean	SD
Familiarization	11	2.43	0.77	11.75	2.09
One-showing	10	2.15	0.60	10.38	1.28
Two-showing	7	2.31	0.92	12.41	0.92

Wulff, J. J. The teaching effectiveness of a filmed mechanical assembly demonstration with supplementary nomenclature training. Air Force Personnel and Training Research Center, Training Aids Research Laboratory, Unpublished Lab. Note TARL-LN-55-8 (mimeo.), 1955.

This experimental comparison was made as a supplement to the study reported in Chapter 10. It was concerned with evaluating the effects of giving nomenclature training on the names of objects shown in a training film (in addition to other familiarization training given prior to showing the film), when the task to be learned by viewing the film involves assembling unfamiliar component parts with unfamiliar names.

The comparison reported here involved an additional group of 11 subjects, selected and tested in the same way as the three groups described in Chapter 10, which was given prior familiarization with nomenclature training. The experimental procedure for this group, requiring a total of 47 minutes per subject, was identical with that for the prior-familiarization group as described in Chapter 10, except for added nomenclature-training features. These consisted of having the experimenter name each part as it was introduced during the first familiarization trial, and then, after the important features of the part had been pointed out, asking the subject to call back the name of the part to the experimenter, who corrected any errors. During the second familiarization trial, the subject was asked to name each part as it was presented, both before and after the experimenter had pointed out the important features of the part, with any errors again corrected by the experimenter. This nomenclature group was significantly superior to the no-familiarization group in terms of the Wilcoxon test for unpaired replicates (one tail) for selection time (P: .01) as well as for selection errors (P: .01) and assembly errors (P: .03) --cf. Chapter 10. (These differences might, of course, be accounted for mainly in terms of the very large training-time differential between the prior-familiarization groups and the no-familiarization group.) The comparisons of primary interest here are for the differences between the familiarization-plus-nomenclature group and the familiarization-without-nomenclature group, shown below.

Mean Test Scores for Each of the Four Criteria

	Prior Familiarization Only	Prior Familiarization Plus Nomenclature
Selection errors	6.8	5.7
Selection time (seconds)	194	137
Assembly errors	22.8	14.9
Assembly time (seconds)	776	635

Though the obtained means favor the nomenclature group, they were shown to be significant only for selection time (P less than .01, one tail) with the Wilcoxon unpaired-replicates test. A more sensitive test which, in view of the substantial differences in means, might have shown some of the other differences to be significant cannot be made with the data now available.

Zuckerman, J. V. Testing with a pre-release filmstrip as a means of predicting factual learning from a training film. Washington, D. C.: U. S. Air Force Human Resources Research Laboratories, HRRL Memo. Report No. 14, November 1951.

Later reports on this study have been published by Zuckerman (1954) and by Smith, Aukes, and Lumsdaine (1954). It was concerned with ascertaining experimentally the extent to which responses of typical students to a preliminary filmstrip, made from a series of "story-board" sketches, can serve as a predictive device to tell how well students would learn specific points of information from a completed film. The sketches were intended to represent the most important visual aspects of the scenes to be photographed in the motion picture, "Flight Capabilities of the F86-A," with the accompanying dialogue or narration indicated for each sketch. Two groups of subjects were compared: Group I viewed a black-and-white film-strip made from about 250 story-board sketches; Group II viewed the completed color motion picture. The same sound track was used for both groups. Both groups were given pre- and post-tests covering in detail the points of information presented in the films. Subjects were 90 Air Force student pilots in advanced single-engine pilot training. The results, based on percentages of correct answers for each of the 26 multiple-choice test items, were computed in terms of: (a) after scores; (b) absolute gain scores (after-minus-before percentages); (c) adjusted absolute gain scores (after-minus-before differences with the "before" base stabilized by lumping the before-test scores for both film and filmstrip group, and basing the gains on the combined before-percentages); and (d) adjusted "relative-gain" scores (obtained by dividing the difference between after-scores and combined before-values by the maximum change possible). The level of learning attained varied widely among the various points of information covered by the films. Pearson product-moment correlation coefficients, computed between the pairs of item percentages for the two treatments, were .80, .71, .94, and .89, respectively, for the four measures. Regardless of which measure is considered, the varying levels of knowledge on specific points obtained from the film are thus predicted fairly well by the levels of knowledge obtained from the filmstrip. This means that tryout data from the preliminary form could have been used profitably for identifying how well various points of information would get across and, hence, as a basis for knowing on which points the presentation would be satisfactory and on which points it should be strengthened. (The value of this kind of application has been demonstrated in a later study by Gropper, Lumsdaine, and Shipman, 1961.) An incidental finding in the present study was that the average level of achievement for the final, polished film was roughly the same as for the crude preliminary version (see also the similar results reported by May and Lumsdaine, 1958, Chapter 2).

BIBLIOGRAPHY AND AUTHOR INDEX

Note: This is a consolidated bibliography of all the references cited by authors of chapters in the book. For explanation of abbreviations or organizations and designation of categories of reports, the reader is referred to the Introduction to the Appendices (pp. 574-579) of the source book, "Teaching Machines and Programmed Learning," by Lumsdaine and Glaser (1960). Unpublished documents, papers, and theses are indicated herein by "U," and publications with limited availability are designated by "L." Annotations are included in citation of reports with limited availability to indicate reports that are superceded by later, more widely available ones. Where two available reports are substantially the same, they are cross-referenced by a "see also" or "see" notation. In cases where a given authorship has more than one publication within a given year, parenthetical letters (a), (b), etc., are used to identify the several publications within that year, which are arranged alphabetically by title.

The underlined numbers at the end of each citation indicate pages of this volume which make reference to this publication. Since this reference list thus also serves as an author index, non-senior co-authors are also cited alphabetically within the bibliography, with references to publications of which they are co-authors. Publications that are chapters in this volume or are abstracted in the Appendix are noted as such, i.e., "Ch. 2 herein" and "Abstr. herein." When the availability of a publication source is limited, but the article is published in full or abstracted in TMPL, this is also indicated in the citation, i.e., "In TMPL," or "TMPL abstr."

Allen, W. H. Research on film use: student participation. Audio-Visual Commun. Rev., 1957, 5, 423-450. (TMPL abs., pp. 581-582)--7.

Allen, W. H. Audio-visual communication research. In C. W. Harris (Ed.), Encyclopedia of educational research. (3rd ed.) New York: Macmillan Co., 1960. Pp. 115-137.--7.

Ammons, R. B. Acquisition of motor skill: I. Quantitative analyses and theoretical formulation. Psychol. Rev., 1947(a), 54, 263-281.--177, 182, 199, 205.

Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. J. exp. Psychol., 1947(b), 37, 393-411.--136, 182, 192, 199.

Ammons, R. B. Rotary pursuit apparatus: I. Survey of variables. Psychol. Bull., 1955, 52, 69-76.--136.

Angell, D., & Lumsdaine, A. A. Stimulus presentation factors relevant to automated teaching. Paper read at meeting of Psychonomic Society, Chicago, September 1960 (a). U (superceded by Ch. 24 herein).--363, 386, 389, 403, 480.

Angell, D., & Lumsdaine, A. A. Prompted and unprompted trials versus prompted trials only in paired-associates learning. Pittsburgh, Pa.: American Institute for Research, 1960(b). L (TMPL abs., p. 582; superceded by Ch. 24 herein).--363, 386, 389, 403.

Angell, D., & Lumsdaine, A. A. The effects of prompting trials and partial-correction procedures on learning by anticipation. San Mateo, Calif.: American Institute for Research. (AFOSR-TN-61-1343), September 1961(a): Abstr. herein.--478, 480, 490, 501, 502.

Angell, D., & Lumsdaine, A. A. Prompted and unprompted trials versus prompted trials only in paired-associates learning. 1961(b): Ch. 24 herein.--229, 330, 331, 349, 363, 386, 475, 480, 481, 482.

Angell, D., & Lumsdaine, A. A. A study of subject-controlled partial cueing in paired-associate learning. San Mateo, Calif.: American Institute for Research (AFOSR-TN-61-1342), September 1961(c): Abstr. herein.--501, 503.

Arps, G. F. Work with knowledge of results versus work without knowledge of results. Psychol. Monogr., 1920, 28, No. 3 (Whole No. 125).--272.

Ash, P. The relative effectiveness of massed versus spaced film presentation. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: US Naval Training Device Center. Office of Naval Research, Tech. Rept. No. SDC 269-7-3, 1949.--205.

Auble, D. Extended tables for the Mann-Whitney statistic. Bull. Inst. Educ. Res. at Ind. Univ., 1953, 1 (2).

Aukes, L. E., see Smith, Aukes, & Lumsdaine.

Baker, Katherine, see Gagné, Baker, & Foster.

Besnard, G. G., see also Briggs & Besnard.

Besnard, G. G., Briggs, L. J., Mursch, G. A., & Walker, E. S. Development of the Subject-Matter Trainer. AFPTRC, Armament Systems Personnel Research Laboratory, Tech. Memo. ASPRL-TM-55-7, March 1955. L (TMPL abs., p. 585; see Briggs, 1958)--401, 402.

Besnard, G. G., Briggs, L. J., & Walker, E. S. The Improved Subject-Matter Trainer. AFPTRC, Armament Systems Personnel Research Laboratory, Tech. Memo. ASPRL-TM-55-11, April 1955. L (TMPL abs., p. 585; see Briggs, 1958)--402.

Briggs, L. J. Two self-instructional devices. Psychol. Rep., 1958, 4, 671-676. (In TMPL, pp. 299-304)--376, 386, 402.

Briggs, L. J. Teaching Machines for training military personnel in maintenance of electronic equipment. In E. H. Galanter (Ed.), Automatic teaching: the state of the art. New York: John Wiley and Sons, Inc., 1959. Pp. 131-146. (TMPL abs., pp. 588-589)--401.

Briggs, L. J. A survey of cueing methods in education and in automated programs. Pittsburgh, Pa.: American Institute for Research (AFOSR-TN-60-286), May 1960. L (TMPL abs., p. 589)--390, 474.

Briggs, L. J. Prompting and confirmation conditions for three learning tasks employing the Subject-Matter Trainer. 1961(a): Ch. 23 herein.--330, 389, 400, 474, 475, 480, 481, 491.

Briggs, L. J. Self-pacing versus automatic pacing in paired-associate learning. 1961(b): Ch. 25 herein.--331, 375, 376.

Briggs, L. J., see also Besnard, Briggs, Mursch, & Walker; Besnard, Briggs, & Walker; Irion & Briggs.

Briggs, L. J., & Besnard, G. G. Experimental procedures for increasing reinforced practice in training Air Force mechanics for an electronic system. In G. Finch & F. Cameron (Eds.), Symposium on Air Force human engineering, personnel, and training research. Washington, D. C.: National Academy of Sciences—National Research Council, Publication No. 445, 1956. Pp. 48-58. (TMPL abs., p. 591)--386, 401.

Briggs, L. J., & DuVall, W. E. Design of two fire control system maintenance training devices. AFPTRC, Technical Report AFPTRC-TR-57-7 (ASTIA Doc. No. 134242), September 1957. (TMPL abs., p. 591)--401.

Briggs, L. J., Plashinski, D., & Jones, D. L. Self-pacing versus automatic pacing of practice on the Subject-Matter Trainer. AFPTRC, Armament Systems Personnel Research Laboratory, Lab. Note ASPRL-LN-55-8, September 1955. U--mimeo. (TMPL abs., pp. 591-592; superceded by Briggs, 1961b; Ch. 25 herein)--399, 401.

Broussard, G. G., see Herman, Broussard, & Todd.

Brown, J. E. Familiarity and novelty of stimulus and response terms in paired associate learning. Master's thesis, Department of Psychology, North Carolina State College, 1960. U--363.

Campbell, V. N., see Goldbeck, Campbell, & Llewellyn.

Carr, H. A. Teaching and learning. J. gen. Psychol., 1930, 37, 182-219.--272.

Carr, W. J. Self-instructional devices: a review of current concepts. Wright-Patterson Air Force Base, Ohio: Aero-Space Medical Laboratory, Wright Air Development Center, WADC Technical Report 59-503, August 1959. (In TMPL, pp. 540-562, under title, "A functional analysis of self-instructional devices.")--330.

Child, I. L. Personality. Annu. Rev. Psychol., 1954, 5.--420.

Clugston, H. A., see Vandell, Davis, & Clugston.

Cook, J. O. Processes underlying the learning of a single paired-associate item. AFPTRC, Maintenance Laboratory, Staff Memorandum, September 1956. U--mimeo. (TMPL abs., p. 601; preliminary version of Cook, 1958; see also Ch. 21 herein)--403.

Cook, J. O. Supplementary report: processes underlying learning a single paired-associate item. J. exp. Psychol., 1958, 56, 455. (TMPL abs., pp. 601-602; see also Ch. 21 herein)--349, 363, 373, 386, 389, 392, 393ff., 403.

Cook, J. O. Research in audio-visual communication. In J. Ball & F. C. Byrnes (Eds.), Research, principles, and practices in visual communication. East Lansing, Michigan: Michigan State University, National Project in Agricultural Communications, 1960(a). Pp. 91-106. (TMPL abs., p. 602)--7, 137.

Cook, J. O. Substitution errors in paired associate learning. Manuscript, North Carolina State College; 1960(b). U--363.

Cook, J. O. From audience participation to paired-associates learning. 1961(a): Ch. 21 herein.--330, 331, 332, 389, 408, 474, 475, 481, 482, 490.

Cook, J. O. Response analysis in paired-associate learning experiments. 1961(b): Ch. 22 herein.--330, 331, 363, 386, 389, 391, 408.

Cook, J. O., see also Kendler, Cook & Kendler; Kendler, Kendler, & Cook.

Cook, J. O., & Kendler, Tracy S. A theoretical model to explain some paired-associate learning data. In G. Finch & F. Cameron (Eds.), Symposium on Air Force human engineering, personnel, and training research. Washington,

D. C.: National Academy of Sciences—National Research Council, 1956. (TMPL abs., pp. 602-604; see also Cook, 1961a: Ch. 21 herein)--349, 351, 358, 361, 389.

Cook, J. O., & Spitzer, M. E. Supplementary report: prompting versus confirmation in paired-associate learning. J. exp. Psychol., 1960, 59, 257-276. (TMPL abs., pp. 604-605)--356, 363, 370, 372, 389, 488.

Cook, T. W. Amount of material and difficulty of the problem solving. II. the disc transfer problem. J. exp. Psychol., 1937, 20, 288-296.--193.

Cotterman, T. E. Task classification: an approach to partially ordering information on human learning. Wright-Patterson Air Force Base, Ohio: Wright Air Development Center, WADC Technical Note 58-374, January 1959.--459.

Coulson, J. E., & Silberman, H. F. Results of an initial experiment in automated teaching. Santa Monica, Calif.: System Development Corp., July 1959. L (In TMPL, pp. 452-468)--489.

Daily, J. T. Development of the Airman Classification Test Battery. U. S. Air Force Human Resources Research Center, Res. Bull. 48-4, November 1948. --451.

Davis, R. A., see Vandell, Davis, & Clugston.

Deese, J. E., see Lazarus, Deese, & Osler.

Detambel, M. H., & Stolurow, L. M. Stimulus sequence and concept learning. J. exp. Psychol., 1956, 51, 34-40.--445.

DuBois, S. Jane, see Weiss, Sheffield, & DuBois.

Duncan, D. B. Multiple range and multiple F tests. Biometrics, 1955, 11, 1-42. --344.

Duva, J. S., & Lumsdaine, A. A. The influence of image quality on the teaching effectiveness of the Kinescope. Washington, D.C.: U. S. Air Force, HFORL, 1956. U.

DuVall, W. E., see Briggs & DuVall.

Eckstrand, G. The influence of a relevant but "unused" cue in training upon transfer in a positive transfer situation. Wright-Patterson Air Force Base, Ohio: Wright Air Development Center, WADC Tech. Rept. 52-79, Part I, April 1952.--142.

Edwards, A. L. Statistical methods for the behavioral sciences. New York: Rinehart and Company, 1954.--380.

Edwards, W. D. The use of statistical decision functions in making practical decisions. In G. Finch & F. Cameron (Eds.), Symposium on Air Force human engineering, personnel, and training research. Washington, D. C.: National Academy of Sciences—National Research Council, 1956. Pp. 115-124. (TMPL abs., pp. 614-615)--251.

Ehrensberger, R. An experimental study of the relative effectiveness of certain forms of emphasis in public speaking. Speech Monogr., 1945, 12, 94-111. --192.

Emeson, D. L., see also Wulff & Emeson.

Emeson, D. L., & Wulff, J. J. The relationship between "what is learned" and "how it's taught." AFPTRC, Maintenance Laboratory, Tech. Memo. ML-TM-57-32, December 1957. L (TMPL abs., pp. 616-618; see Wulff and Emeson, 1961: Ch. 30 herein)--457.

Evans, J. L., Glaser, R., & Homme, L. E. A preliminary investigation of variation in the properties of verbal learning sequences of the "teaching machine" types. Paper read at the Eastern Psychological Association Convention, Atlantic City, New Jersey, April 1959. (In TMPL, pp. 446-451)--488.

Fairbanks, G., Guttman, N., & Miron, M. S. The effects of time compression upon the auditory comprehension of spoken messages. AFPTRC Technical Report AFPTRC-TN-57-73 (ASTIA Doc. No. 131425), June 1957: Abstr. herein. --502, 504.

Faison, E. W. J., Rose, N., & Podell, J. E. The effects of rest pauses during training film instruction. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-44, 1955(a). U--mimeo: Abstr. herein.--205, 441, 502, 506.

Faison, E. W. J., Rose, N., & Podell, J. E. A technique for measuring observable audience reactions to training films. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-45, 1955(b). U--mimeo: Abstr. herein.--205, 506.

Falsinger, J. M., see Gladstone, Yamaguchi, Hull, & Falsinger.

Foster, Harriet, see Gagné, Baker, & Foster; Gagné & Foster.

Fine, B. J., see Weiss & Fine.

Fry, E. B. A study of teaching machine response modes. In A. A. Lumsdaine & R. Glaser (Eds.), Teaching machines and programmed learning: a source book. Washington, D.C.: National Education Association, 1960. Pp. 469-474.--331.

Gagné, R. M. The effect of sequence of presentation of similar items on the learning of paired associates. J. exp. Psychol., 1950, 40, 61-73.

Gagné, R. M., Baker, Katherine E., & Foster, Harriet. Transfer of discrimination training to a motor task. J. exp. Psychol., 1950, 40 (3), 314-328.

Gagné, R. M., & Foster, Harriet. A study of transfer in a motor task with varying display control relationships. U. S. Naval Medical Research Laboratory, Lab. Report 316-1-3, 1950. (see Also Gagné, Baker, & Foster)--157.

Gates, A. I. Recitation as a factor in memorizing. Arch. Psychol., 1917, 7 (40). (TMPL abs., p. 629)--3, 4, 6, 137, 417.

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. Psychol. Rev., 1940, 47, 196-229.--347, 372.

Gibson, E. J. Intralist generalization as a factor in verbal learning. J. exp. Psychol., 1942, 30, 185-200.

Gilbert, T. F. An early approximation to principles of programming continuous discourse, self-instructional materials. A report to Bell Telephone Laboratories, September 1958. U--mimeo. (TMPL abs., pp. 630-634).

Gladstone, A. I., see also Lumsdaine & Gladstone

Gladstone, A. I., Hamaguchi, H. G., Hull, C. L., & Falsinger, J. M. Some functional relationships of reaction potential (SER) and related phenomena. J. exp. Psychol., 1947, 37, 510-526.--193.

Glaser, R. Some research problems in automated instruction: instructional programming and subject matter structure. In J. E. Coulson (Ed.), Programmed learning and computer-based instruction. New York: John Wiley and Sons, Inc., 1962.--494.

Glaser, R., see also Evans, Glaser, & Homme; Lumsdaine & Glaser.

Goldbeck, R. A. The effect of response mode and learning material difficulty on automated instruction. Santa Barbara, Calif.: American Institute for Research, Tech. Report No. 1, Contract No. Nonr-3077(00), September 1960. --487, 488.

Goldbeck, R. A., Campbell, V. N., & Llewellyn, Joan E. Further experimental evidence on response modes in automated instruction. Santa Barbara, Calif.:

American Institute for Research, Tech. Report No. 3, Contract No. Nonr-3077(00), December 1960.--487.

Grier, J. B. Stimulus-response asynchronism in paired associate learning. Master's thesis, Department of Psychology, North Carolina State College, 1960. U--363.

Gropper, G. L., Lumsdaine, A. A., & Shipman, Virginia. Studies in televised instruction. Report No. 1: improvement of televised instruction based on student responses to achievement tests. Pittsburgh, Pa.: American Institute for Research, March 1961: Abstr. herein.--502, 508, 527.

Grosslight, J. H., see Kale, Grosslight & McIntyre.

Guthrie, E. R. The psychology of learning. New York: Harper and Bros., Inc., 1935. (Revised edition, 1952)

Guthrie, P. M., & Lumsdaine, A. A. Some effects of graduated partial cueing on the learning of paired associates. San Mateo, Calif.: American Institute for Research (AFOSR-TN-61-1341), December 1961: Abstr. herein.--480, 501, 509.

Guttman, N., see Fairbanks, Guttman & Miron.

Hadsell, R. S., see Lumsdaine, May & Hadsell.

Harby, S. F. Comparison of mental practice and physical practice in the learning of physical skills. (Pennsylvania State University Instructional Film Research Program). Port Washington, N.Y.: U. S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-27, 1952.--419.

Herman, D. T., Broussard, I. G., & Todd, H. R. Inter-trial interval and the rate of learning serial order picture material. J. gen. Psychol., 1951, 45, 245-254.--336.

Hilgard, E. R. Theories of human learning and problems of training. In Symposium on psychology of learning basic to military problems. Washington, D. C.: Committee on Human Resources, Research and Development Board, May 1953.--417.

Hilgard, E. R. Theories of learning. (2nd ed.) New York: Appleton-Century-Crofts, Inc., 1956.

Hilgard, E. R., & Marquis, D. G. Conditioning and learning. New York: Appleton-Century-Crofts, Inc., 1940. (2nd ed., 1961)--19.

Hirsch, R. S. The effects of knowledge of test results on learning of meaningful material. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: U. S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-30, 1952.--491.

Hoban, C. F., Jr., & van Ormer, E. B. Instructional film research, 1918-1950. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: U. S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-19, 1950.

Hoehn, A. J., see also Sheffield, Margolius, & Hoehn.

Hoehn, A. J., & Lumsdaine, A. A. Design and use of job aids for communicating technical information. AFPTRC, Technical Report AFPTRC-TR-58-7 (ASTIA Doc. No. AD 152109), January 1958. (TMPL abs., p. 639)--iv.

Holland, J. G. Program design and use. A paper read at the American Psychological Association Convention, Chicago, Ill., 1960. U--mimeo.--487, 488.

Holland, J. G., see also Skinner & Holland.

Hollingworth, H. L. General laws of redintegration. J. gen. Psychol., 1928, 1, 79-90.--17.

Hollis, A. P. The effectiveness of the motion picture, demonstration by the teacher, and oral instruction in teaching cooking. In F. N. Freeman (Ed.), Visual education. Chicago: University of Chicago Press, 1924. Pp. 339-341.--155.

Homme, L. E., see Evans, Glaser, & Homme.

Hovland, C. I. The generalization of conditioned responses. I. the sensory generalization of conditioned responses with varying frequencies of tone. J. gen. Psychol., 1937(a), 17, 135-148.--156.

Hovland, C. I. The generalization of conditioned responses. II. the sensory generalization of conditioned responses with varying intensities of tone. J. genet. Psychol., 1937(b), 51, 279-291.--156.

Hovland, C. I. Experimental studies in rote-learning theory. I. reminiscence following learning by massed and by distributed practice. J. exp. Psychol., 1938, 22, 201-224.--205.

Hovland, C. I. Experimental studies in rote-learning theory. VII. distribution of practice with varying length of lists. J. exp. Psychol., 1940, 27, 271-284. --193.

Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: John Wiley and Sons, Inc., 1951. Pp. 613-689.--3, 5, 272, 287.

Hovland, C. I., et al. The order of presentation in persuasion. New Haven: Yale University Press, 1957.--500.

Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. Experiments on mass communication. Princeton, N. J.: Princeton University Press, 1949--5, 6, 139, 198, 213, 227, 271, 272, 280, 281, 282, 286, 296, 313, 320, 321, 332, 418, 419, 427, 429, 472, 479, 489, 498, 500.

Hull, C. L. Principles of behavior. New York: Appleton-Century-Crofts, Inc., 1943.--156.

Hull, C. L. Essentials of behavior. New Haven: Yale University Press, 1951.--215.

Hull, C. L., see also Gladstone, Yamaguchi, Hull, & Falsinger.

Hurlock, E. B. An evaluation of certain incentives used in school work. J. educ. Psychol., 1925, 16, 145-159.--512.

Irion, A. L., see also McGeoch & Irion.

Irion, A. L., & Briggs, L. J. Learning task and mode of operation variables in use of the Subject-Matter Trainer. AFPTRC, Tech. Rept. AFPTRC-TR-57-8 (ASTIA Doc. No. AD 134252), October 1957. (TMPL abs., pp. 642-643; see Briggs, 1961a: Ch. 23 herein)--349, 375, 401.

Jackson, R. Learning from kinescopes and films. Port Washington, N.Y.: U. S. Naval Special Device Center, Office of Naval Research, Tech. Rept. No. SDC 20-TV-1, 1952.

Jacobs, M., see Zuckerman & Jacobs.

Janis, I. L., & King, B. T. The influence of role playing on opinion change. J. abnorm. Psychol., 1954, 49, 211-218.--418.

Jaspen, N. Contribution of especially designed sound motion pictures to the learning of skills: assembly of the breech block of the 40mm. antiaircraft gun. Doctoral dissertation, Pennsylvania State University, 1949. U (see also Jaspen, 1950a,b)--155, 193, 196.

Jaspen, N. Effects on training of experimental film, study I. verbalization, rate of development, nomenclature, errors, "how-it-works," repetition. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: U. S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-17, October 1950(a).--177, 193.

Jaspen, N. Effects on training of experimental film variables, study II. verbalization, "how-it-works," nomenclature, audience participation, and succinct treatment. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: U. S. Naval Training Device Center, Office of Naval Research. Tech. Rept. No. SDC 269-7-11, March 1950(b).--420.

Jones, D. L., see Briggs, Plashinski, & Jones.

Judd, C. H. Practice without knowledge of results. Psychol. Monogr., 1905, 7, No. 1 (Whole No. 29), 185-198.--272.

Kaess, W., & Zeaman, D. Positive and negative knowledge of results on a Pressey-type punchboard. J. exp. Psychol., 1960, 60, 12-17.--397.

Kale, S. V., Grosslight, J. H., & McIntyre, C. J. Exploratory studies in the use of pictures and sound for teaching foreign language vocabulary. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: U. S. Naval Training Devices Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-53, August 1955.--338.

Kanner, J. H., & McClure, A. H. Varied versus identical repetition in filmed instruction on micrometer reading. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL LN-56-11, April 1956. U--mimeo: Abstr. herein.--362, 495, 502, 510.

Kanner, J. H., & Sulzer, R. L. Overt and covert rehearsal of 50% versus 100% of the material in film learning. AFPTRC, Training Aids Research Laboratory, Staff Memo. TARL 56-12, 1956. U--mimeo. (superceded by Ch. 28 herein) --427.

Kanner, J. H., & Sulzer, R. L. Overt and covert rehearsal of fifty per cent versus one hundred per cent of the material in film learning. 1961: Ch. 28 herein. --139, 332, 478, 486, 487, 488, 489, 490.

Kendler, H. H., see Kendler, Cook, & Kendler; Kendler, Kendler, & Cook.

Kendler, Tracy S., see also Cook & Kendler.

Kendler, Tracy S., Cook, J. O., & Kendler, H. H. An investigation of the interacting effects of repetition and audience participation on learning from films. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-56-2, January 1956, U--mimeo. (see Cook, 1961a: Ch. 21 herein)--352, 365.

Kendler, Tracy S., Kendler, H. H., & Cook, J. O. Implications of learning theory for the design of audio-visual aids. Washington, D.C.: U. S. Air Force Human Resources Research Laboratories, HRRL Memo. Report No. 12(a), November 1951.--352.

Kendler, Tracy S., Kendler, H. H., & Cook, J. O. Effect of opportunity and instructions to practice during a training film on initial recall and retention. AFPTRC, Training Aids Research Laboratory, Staff Research Memo., 29 March 1954. U--mimeo. (see Cook, 1961a: Ch. 21 herein)--354, 366.

Kimble, G. A. An experimental test of a two-factor theory of inhibition. J. exp. Psychol., 1949, 39, 15-23.--184.

Kimble, G. A. The effect of praise, reproof, and no incentive on the value of a training film. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-7, 1955(a). U--mimeo. Abstr. herein.--137, 502, 512, 514, 515.

Kimble, G. A. The value of an interpolated test in increasing the effectiveness of a training film. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-10, December 1955(b). U--mimeo: Abstr. herein.--479, 502, 512, 514, 518.

Kimble, G. A., & Wulff, J. J. The effect of "response guidance" on the value of audience participation in training film instruction. Washington, D. C.: U. S. Air Force Human Factors Operations Research Laboratories, HFORL Report No. 34, 1953(a). L (TMPL abs., pp. 646-647; see also Ch. 15 herein) --228, 239.

Kimble, G. A., & Wulff, J. J. "Response guidance" as a factor in the value of audience participation in training film instruction. Washington, D. C.: U.S. Air Force Human Factors Operations Research Laboratories, HFORL Report No. 36, 1953(b). (Same as 1953a, above, except lacks technical appendix; see also Ch. 15 herein)--217, 228, 239.

Kimble, G. A., & Wulff, J. J. The teaching effectiveness of instruction in reading a scale as a function of the relative amounts of problem solving practice and demonstration examples used in training. AFPTRC, Training Aids Research Laboratory, Staff Research Memo. 54-22, March 1954. U--mimeo. (superceded by Ch. 16 herein)--227.

Kimble, G. A., & Wulff, J. J. "Response guidance" as a factor in the value of student participation in film instruction. 1961(a): Ch. 15 herein.--137, 138, 263, 287, 441, 474, 475, 481, 487, 490, 514, 515.

Kimble, G. A., & Wulff, J. J. The effectiveness of instruction in reading a scale as influenced by the relative amounts of demonstration and problem-solving practice. 1961(b): Ch. 16 herein.--6, 137, 420.

King, B. T., see Janis & King.

Klaus, D. J. The art of auto-instructional programming. Audio-Visual Commun. Rev., 1961, 9, 130-142.--474, 482.

Klier, S., see Miller & Klier.

Kopstein, F. F. An experimental exploration of the interaction between teaching and learning behavior. Doctoral dissertation, University of Illinois, 1961. U--336, 350.

Kopstein, F. F., see also Lumsdaine, Sulzer, & Kopstein; Sidowski, Kopstein, & Shillestad; Sulzer, Lumsdaine, & Kopstein.

Kopstein, F. F., & Roshal, S. M. Learning foreign vocabulary from pictures versus words. Amer. Psychologist, 1954, 9, 407-408. (Abstract; see Ch. 20 herein)--335, 338.

Kopstein, F. F., & Roshal, S. M. Method of presenting word pairs as a factor in foreign vocabulary learning. Amer. Psychologist, 1955, 10, 354. (Abstract; see Ch. 20 herein)--335, 369.

Kopstein, F. F., & Roshal, S. M. Verbal learning efficiency as influenced by the manipulation of representational response processes: pictorial-verbal and temporal contiguity factors. 1961: Ch. 20 herein.--6, 138, 329, 475, 482, 484.

Kraeling, Doris G., see Wulff & Kraeling; Wulff, Sheffield, & Kraeling.

Krueger, Louise O. The relative effect of interspersing a recall at different stages of learning. Arch. Psychol., 1930, 18 (114), 15-25.--420.

Krueger, W. C. F. The effect of overlearning on retention. J. exp. Psychol., 1929, 12, 71-78.--193, 197.

Krueger, W. C. F. Further studies in overlearning. J. exp. Psychol., 1930(a), 13, 152-163.--483.

Krueger, W. C. F. The optimal effect of a recall during learning. Arch. Psychol., 1930(b), 18 (114), 26-34.--206, 420, 483.

Kurtz, K. H. Transfer of discrimination learning with complex stimuli. Doctoral dissertation, Yale University, 1953. U-142.

Kurtz, K. H. Discrimination of complex stimuli: the relationshop of training and test stimuli in transfer of discrimination. J. exp. Psychol., 1955, 50, 283-292.--445, 454.

Lazarus, M., Deese, J. E., & Osler, Sonia F. The effects of psychological stress upon performance. Psychol. Bull., 1952, 49, 293-217.--182.

Levine, S. The role of motivation in the effects of "active review" on learning from a factual film. Amer. Psychologist, 1953, 8, 338-389. (Abstract; a preliminary report on the following study.)

Levine, S. Experiment on further factors influencing the effects of audience participation. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-43, December 1955. U--mimeo. (superceded by Ch. 19 herein; see also preceding reference)--295.

Levine, S., see also Maccoby, Michael, & Levins; Miller & Levine; Miller, Levine, & Sternberg.

Liberman, A. M. Studies in the perception of synthetic speech. Paper read at annual meeting of American Psychological Association, Cleveland, Ohio, September 1954. U--355.

Lindley, R. H., see Shettel & Lindley.

Lindquist, E. F. Statistical analysis in educational research. Boston: Houghton Mifflin Company, 1940.--166.

Llewellyn, Joan E., see Goldbeck, Campbell, & Llewellyn.

Lockhart, J. A. The value of the motion picture as an instructional device in learning a motor skill. Research Quarterly of the Amer. Ass'n. for Health, Recreation, and Physical Ed., 1944, 15, 181-187.--155.

Lumsdaine, A. A. Ease of learning with pictorial and verbal symbols. Doctoral dissertation, Stanford University, 1949. U (see Lumsdaine, 1958)--329, 336, 347, 484.

Lumsdaine, A. A. Ease of learning with pictorial and verbal symbols. Stan. Univ. Bull., 30 November 1950, 8th series, No. 51, 222 (abstract of Lumsdaine, 1949)--329, 336, 347, 484.

Lumsdaine, A. A. Audio-visual research in the U. S. Air Force. Audio-Visual Commun. Rev., 1953, 1, 76-90.--7.

Lumsdaine, A. A. Cue and response functions of pictures and words. In M. A. May & A. A. Lumsdaine (Eds.), Learning from films. New Haven: Yale University Press, 1958. Pp. 123-149.--329, 336, 347, 484.

Lumsdaine, A. A. Partial and more complete automation of teaching in group and individual learning situations. In E. H. Galanter (Ed.), Automatic teaching: the state of the art. New York: John Wiley and Sons, Inc., 1959(a). (TMPL abs., pp. 650-651)--ii, 2, 5, 7, 11.

Lumsdaine, A. A. Some issues concerning devices and programs for automated learning. 1959(b). (In TMPL, pp. 517-539)--330, 398, 476, 482, 483.

Lumsdaine, A. A. Teaching machines and self-instructional materials. Audio-Visual Commun. Rev., 1959(c), 7, 163-181. (In TMPL under title, "Teaching machines: an introductory overview," pp. 5-22)--iv, 252.

Lumsdaine, A. A. The analysis of student response as a factor in instruction. 1961: Ch. 1 herein.

Lumsdaine, A. A. Instruments and media of instruction. In N. L. Gage (Ed.), Handbook of research on teaching. Washington, D. C.: American Educational Research Association, National Education Association, 1962(a). (in press)--251.

Lumsdaine, A. A. Some theoretical and practical problems in programmed instruction. In J. E. Coulson (Ed.), Programmed learning and computer-based instruction. New York: John Wiley and Sons, Inc., 1962(b). (in press)—471, 491, 496.

Lumsdaine, A. A., see also Angell & Lumsdaine; Duva & Lumsdaine; Gropper, Lumsdaine, & Shipman; Guthrie & Lumsdaine; Hoehn & Lumsdaine; Hovland, Lumsdaine & Sheffield; May & Lumsdaine; Smith, Aukes, & Lumsdaine; Sulzer, Lumsdaine, & Kopstein.

Lumsdaine, A. A., & Gladstone, A. I. Overt practice and audio-visual embellishments. In M. A. May & A. A. Lumsdaine (Eds.), Learning from films. New Haven: Yale University Press, 1958. Pp. 58-71. (TMPL abs., p. 655)--6.

Lumsdaine, A. A. & Glaser, R. (Eds.) Teaching machines and programmed learning: a source book. Washington, D. C.: National Education Association, 1960.--iv, 401, 457, 471, 472, 501.

Lumsdaine, A. A., May, M. A., & Hadsell, R. S. Do "motivation" and "participation" questions increase learning? Educational Screen, May 1947. (see also Lumsdaine, May, & Hadsell, 1958)--193, 206, 271, 272, 298, 306, 314.

Lumsdaine, A. A., May, M. A., & Hadsell, R. S. Questions spliced into a film for motivation and pupil participation. In M. A. May and A. A. Lumsdaine (Eds.), Learning from films. New Haven: Yale University Press, 1958. Pp. 72-83. (also referenced as Yale Motion Picture Research Project, 1947) --192, 193, 271, 272, 298, 306, 314.

Lumsdaine, A. A., & Sulzer, R. L. The influence of simple animation techniques on the value of a training film. Washington, D. C.: U. S. Air Force Human Resources Research Laboratories, HRRL Report No. 24, 1951 (see Lumsdaine, Sulzer, and Kopstein, 1961: Ch. 17 herein)--241.

Lumsdaine, A. A., Sulzer, R. L., & Kopstein, F. F. The effect of animation cues and repetition of examples on learning from an instructional film. 1961: Ch. 17 herein.--138, 193, 196, 313, 474, 477, 479, 487.

Maccoby, N., see Margolius & Maccoby; Margolius, Sheffield, & Maccoby; Michael & Maccoby; Sheffield & Maccoby; Weiss, Sheffield & Maccoby.

Maccoby, N., Michael, D. N., & Levine, S. Further studies of student participation procedures in film instruction: review and preview covert practice, and motivational interactions. 1961: Ch. 19 herein.--6, 139, 479, 489, 507.

Maccoby, N., & Sheffield, F. D. Summary and implications of some studies of distribution of demonstration and practice in learning serial manual tasks. AFPTRC, Training Aids Research Laboratory, TARL Lab. Note 56-20, March 1956. U--mimeo. (superceded by Maccoby and Sheffield, 1958 and 1961)--77.

Maccoby, N., & Sheffield, F. D. Theory and experimental research on the teaching of complex sequential procedures by alternate demonstration and practice. AFPTRC, Maintenance Laboratory, Tech. Memo. ML-TM-57-31, December 1957. (superceded by Maccoby and Sheffield, 1958 and 1961)--77.

Maccoby, N., & Sheffield, F. D. Theory and experimental research on the teaching of complex sequential procedures by alternate demonstration and practice. In G. Finch & F. Cameron (Eds.), Symposium on Air Force human engineering, personnel, and training research. Washington, D. C.: National Academy of Sciences—National Research Council, Publication No. 516, 1958. (see Ch. 5 herein)--12, 77.

Maccoby, N., & Sheffield, F. D. Combining practice with demonstration in teaching complex sequences: summary and interpretation. 1961: Ch. 5 herein.--6, 330, 386, 407, 474, 475, 476, 480, 483, 492.

Margolius, G. J., see also Sheffield, Margolius, & Hoehn.

Margolius, G. J., & Maccoby, N. Methods of combining practice with filmed demonstration in teaching complex response sequences. Paper read at the annual convention of the American Psychological Association. Amer. Psychologist, 1956, 11, 444. (Abstract) (see Margolius & Sheffield, Ch. 3 herein)

Margolius, G. J., & Sheffield, F. D. Optimum methods of combining practice with filmed demonstration in teaching complex response sequences: serial learning of a mechanical-assembly task. 1961: Ch. 3 herein.--474, 480, 492.

Margolius, G. J., Sheffield, F. D., & Maccoby, N. Repetitive versus continuous demonstration and practice in the learning of a serial mechanical assembly task I. AFPTRC, Maintenance Laboratory, Tech. Memo. ML-TM-57-11, September 1957(a). (see Ch. 6 herein)--87, 90.

Margolius, G. J., Sheffield, F. D., & Maccoby, N. Repetitive versus continuous demonstration and practice in the learning of a serial mechanical assembly task II. AFPTRC, Maintenance Laboratory, Tech. Memo. ML-TM-57-12, September 1957(b). (see Ch. 6 herein)--87, 96.

Margolius, G. J., Sheffield, F. D., & Maccoby, N. Timing of demonstration and overt practice as a function of task organization. AFPTRC, Maintenance Laboratory, Tech. Memo. ML-TM-57-13, September 1957(c). (see Ch. 7 herein)--101.

Margolius, G. J., Sheffield, F. D., & Maccoby, N. Repetitive versus consecutive demonstration and practice in the learning of a serial mechanical-assembly task. 1961(a): Ch. 6 herein.--6, 38, 43, 493, 496.

Margolius, G. J., Sheffield, F. D., & Maccoby, N. Timing of demonstration and overt practice as a function of task organization. 1961(b): Ch. 7 herein.--6, 38, 493.

Margolius, G. J., Sheffield, F. D., & Robinson, Elinor R. Optimum methods of combining practice with filmed demonstration in teaching complex response sequences. I. serial learning of a mechanical assembly task. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-56-15, March 1956. U--mimeo. (superceded by Margolius & Sheffield: Ch. 3 herein)--33.

Marquis, D. G., see Hilgard & Marquis.

May, M. A. The psychology of learning from demonstration films. J. educ. Psychol., 1946, 37, 1-12.--129, 156, 157, 178, 182, 477, 500.

May, M. A. Verbal responses to demonstrational films. In M. A. May & A. A. Lumsdaine (Eds.), Learning from films. New Haven: Yale University Press, 1958.--129, 157, 477, 500.

May, M. A., see also Lumsdaine, May, & Hadsell.

May, M. A., & Lumsdaine, A. A. (Eds.) Learning from films. New Haven: Yale University Press, 1958.--139, 206, 227, 297, 311, 479, 524, 527.

Mayer, Sylvia R., & Westfield, R. L. A field tryout of a teaching machine for training in SAGE operations. Air Force Cambridge Research Center, Operational Applications Laboratory, Tech. Memo. OAL-TM-58-16, October 1958. L (TMPL abs., p. 658)--401.

McClure, A. H., see Kanner & McClure.

McClusky, F. D., & McClusky, H. Y. Comparison of motion pictures, slides, stereographs, and demonstration as a means of teaching how to make a reed mat and a pasteboard box. In F. N. Freeman (Ed.), Visual education. Chicago: University of Chicago Press, 1924. Pp. 310-334.--155.

McClusky, H. Y., see McClusky & McClusky.

McGeoch, J. A., & Irion, A. L. The psychology of human learning. (2nd ed.) New York: Longmans, Green and Company, 1952.--418.

McGuire, W. J. A preliminary study of the use of an instructional film in teaching a motor skill. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-57, December 1955(a). U--mimeo. (superceded by Ch. 12 herein)--177.

McGuire, W. J. The relative efficacy of overt and covert trainee participation with different speed of instruction. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-46, December 1955(b). U--mimeo. (superceded by Ch. 27 herein)--417.

McGuire, W. J. Studies of factors influencing the teaching effectiveness of audio-visual aids. I. length of the film. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-9, December 1955(c). U--mimeo. (superceded by Ch. 13 herein)--187.

McGuire, W. J. Studies of factors influencing the effectiveness of demonstrational films for teaching a motor skill: II. slow motion; added narration; distributed showing. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-50, December 1955(d). U--mimeo. (superceded by Ch. 13 herein)--187.

McGuire, W. J. Studies of factors influencing the effectiveness of demonstrational films for teaching a motor skill: III. serial position and proximity to reward within the film. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-51, December 1955(e). U-mimeo. (superceded by Ch. 14 herein)--209.

McGuire, W. J. Audience participation and audio-visual instruction: overt-covert responding and rate of presentation. 1961(a): Ch. 27 herein.--331, 332, 483, 486, 488.

McGuire, W. J. Effects of serial position and proximity to "reward" within a demonstrational film. 1961(b): Ch. 14 herein.--135, 137, 331, 411, 478.

McGuire, W. J. Interpolated motivational statements within a programmed series of instructions as a distribution-of-practice factor. 1961(c): Ch. 26 herein.--331, 478.

McGuire, W. J. Some deleterious effects on a perceptual-motor skill produced by an instructional film: massing effects, interference, and anxiety. 1961(d): Ch. 12 herein.--135, 331, 478.

McGuire, W. J. Some factors influencing the effectiveness of demonstrational films: repetition of instructions, slow motion, distribution of showings, and explanatory narration. 1961(e): Ch. 13 herein.--135, 136, 137, 138, 252, 331, 478, 479, 495.

McIntyre, C. J., see Kale, Grosslight, & McIntyre.

McTavish, C. L. Effect of repetitive film showings on learning. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: U. S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-12, November 1949.--138, 193, 197, 262, 524.

Melton, A. W. The science of learning and the technology of educational methods. Harvard Educational Rev., 1959, 29, 96-106. (TMPL abs., pp. 658-660)--497, 498, 499.

Meyer, Susan R. Report on the initial test of a junior high-school vocabulary program. 1960. (In TMPL, pp. 229-246)--390, 478.

Michael, D. H., see also Maccoby, Michael, & Levine.

Michael, D. H. (and Maccoby, N.) Some factors influencing the effects of audience participation on learning from a factual film. U. S. Air Force Human Resources Research Laboratories, HRRL Memo. Report 13A, December 1951. L. (see Ch. 18 herein)--138, 271.

Michael, D. N., & Maccoby, N. Factors influencing verbal learning from films under varying conditions of audience participation. J. exp. Psychol., 1953, 46, 411-418. (see also Ch. 18 herein)--227, 271, 279, 419.

Michael, D. N., & Maccoby, N. A further study of the use of "audience participation" procedures in film instruction. AFPTRC, Training Aids Research Laboratory, Staff Research Memo., 25 March 1954. U--mimeo. (superceded by Ch. 19 herein)--295.

Michael, D. N., & Maccoby, N. Factors influencing the effects of student participation on verbal learning from films: motivating versus practice effects, "feedback," and overt versus covert responding. 1961: Ch. 18 herein.--6, 138, 139, 320, 331, 332, 419, 427, 428, 440, 441, 478, 479, 486, 487, 488, 489, 490, 518.

Miller, G. A. The magical number seven, plus or minus two. Psychol. Rev., 1956, 63, 81-97.--198.

Miller, J., & Klier, S. The effect of interpolated quizzes on learning audio-visual material. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-56-8, January 1956(a). U--mimeo. Abstr. herein.--479, 502, 517.

Miller, J., & Klier, S. The effect on active rehearsal types of review of massed and spaced review techniques. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-56-6, January 1956(b). U--mimeo: Abstr. herein.--501, 518.

Miller, J., & Levine, S. A study of the effects of different types of review and of "structuring" sub-titles on the amount learned from a training film. Washington, D. C.: U. S. Air Force Human Resources Research Laboratories, HRRL Memo. Report No. 17, March 1952: Abstr. herein.--193, 205, 493, 495, 501, 507, 518, 519, 521.

Miller, J., Levine, S., & Sternberg, J. The effects of different kinds of review and of sub-titling on learning from a training film. AFPTRC, Training Aids Research Laboratory, Staff Research Memo., June 1954. U--mimeo: Abstr. herein.--501, 518, 521, 522.

Miller, J., Levine, S., & Sternberg, J. Extension to a new subject matter of the findings on the effects of different kinds of review on learning from a training film. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-56-4, January 1956. U--mimeo: Abstr. herein.--495, 501, 518, 522.

Miller, N. E., et al. Graphic communication and the crisis in education. Washington, D. C.: National Education Association, 1957. (Special issue, Audio-Visual Commun. Rev., 1957, 5 (3); TMPL abs., pp. 665-666)--137, 352.

Miron, M. S., see Fairbanks, Guttman, & Miron.

Moll, K. R., see Nelson & Moll.

Murdock, B. B., Jr. Intralist generalization in paired-associate learning. Psychol. Rev., 1958, 60, 306-314.--372.

Murdock, B. B., Jr. A reply to Battig, Gibson and Runquist. Psychol. Rev., 1959, 66, 345-346.

Mursch, G. A., see Besnard, Briggs, Mursch, & Walker.

Nelson, H. E., & Moll, K. R. Comparison of the audio and video elements of instructional films. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: U. S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-18, November 1950.--203.

Neu, D. M. The effect of attention-gaining devices on film-mediated learning. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: U. S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-9, 1950.--198.

Newell, A., Shaw, J. C., & Simon, H. A. Elements of a theory of human problem solving. Psychol. Rev., 1958, 65, 151-166.--487.

Newman, S. E. Student versus instructor design of study methods. J. educ. Psychol., 1957, 48 (6), 328-333. (TMPL abs., pp. 667-668)--333, 336.

Newman, S. E. A selective mediation model of paired associate learning. Paper read at meeting of Psychonomic Society, Chicago, Illinois, September 1960. U--363, 364.

Newman, S. E., see also Saltz & Newman.

Northrop, D. S. Effects on learning of the prominence of organizational outlines in instructional films. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: U. S. Naval Research, Tech. Rept. No. SDC 269-7-33, 1952.--493.

Norton, D. An empirical investigation of some effects of non-normality and heterogeneity on the F-distribution. Doctoral dissertation, State University of Iowa, 1952. U--423.

Osgood, C. E. Method and theory in experimental psychology. New York: Oxford University Press, 1953.--445.

Osler, Sonia F., see Lazarus, Deese, & Osler.

Pan, S. The influence of context upon learning and recall. J. exp. Psychol., 1926, 9, 468-491.--475.

Pask, G. Electronic keyboard teaching machines. Education and Commerce, 1958, 24, 16-26. (In TMPL, pp. 336-348).

Peak, Helen. Negative practice and theories of learning. Psychol. Rev., 1941, 48, 316-336.--192.

Peatman, J. G. Description and sampling statistics. New York: Harper and Brothers Co., 1947.--163.

Perry, H. M. The relative efficiency of actual and "imaginary" practice in five selected tasks. Arch. Psychol., N.Y., 1939, 34 (243).--419.

Plashinski, D., see Briggs, Plashinski, & Jones.

Podell, J. E., see Faison, Rose, & Podell.

Porter, D. A. Some effects of year-long teaching machine instruction. In E. H. Galanter (Ed.), Automatic teaching: the state of the art. New York: John Wiley and Sons, Inc., 1959. Pp. 85-90. (TMPL abs., pp. 673-674)--390.

Pressey, S. L. A machine for automatic teaching of drill material. Sch. & Soc., 1927, 25 (645), 549-552. (In TMPL, pp. 42-46)--386.

Rimland, B. Effectiveness of several methods of repetition of films. (Pennsylvania State University Instructional Film Research Program). Port Washington, N. Y.: U. S. Naval Special Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-45, May 1955.--495.

Robinson, Elinor R., see Margolius, Sheffield, & Robinson.

Roe, A., et al. Automated teaching methods using linear programs. Los Angeles, Calif.: University of California, Department of Engineering, Report No. 60-105, December 1960(a).--487, 488.

Roe, A., et al. A pilot study--automated learning research project. Los Angeles, Calif.: University of California, Department of Engineering, Report No. 60-53, May 1960(b).--487.

Rolfe, E. C. The effectiveness of the film and demonstration in teaching physics. In F. N. Freeman (Ed.), Visual education. Chicago: University of Chicago Press, 1924. Pp. 335-338.

Romney, A. K. Effects of selected "set"-inducing instructions upon learning from multiple film showings. Manuscript, Boston University, September 1952. U: Abstr. herein.--479, 502, 518, 522.

Rose, N., see Faison, Rose, & Podell.

Roshal, S. M. Effects of learner representation in film-mediated perceptual-motor learning. (Pennsylvania State University Instructional Film Research Program.) Port Washington, N. Y.: U. S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-5, 1949(a); (superceded by Ch. 11 herein)--129, 155, 160, 163, 177, 336.

Roshal, S. M. The effectiveness of film-mediated perceptual-motor learning with varying presentations of the task. Doctoral dissertation, Pennsylvania State University, 1949(b). U. (see Ch. 11 herein)--129, 155, 160, 163, 177, 336.

Roshal, S. M. Film-mediated learning with varying representations of the task: viewing angle, portrayal of demonstration, motion, and student participation. 1961: Ch. 11 herein.--135, 420, 477, 485, 487, 495.

Roshal, S. M., see also Kopstein & Roshal.

Rubino, K. J. Comparative learning speed-attainment level of a swing kip on horizontal bars. Doctoral dissertation, University of California at Los Angeles, 1961. U--137.

Rubin-Rabson, Grace. Mental and keyboard overlearning in memorizing piano music. J. Musicol., 1941(a), 3, 33-40.--418, 419.

Rubin-Rabson, Grace. Studies in the psychology of memorizing piano music: VI. a comparison of two forms of mental rehearsal and keyboard overlearning. J. educ. Psychol., 1941(b), 32, 593-602.--418, 419.

Runquist, W. N., see Underwood, Runquist, & Schulz.

Saltz, E., & Newman, S. E. Relationship between language variables and decision making in an assembly task. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-56-22, 1956. U--mimeo. (an earlier version of Saltz & Newman, 1960)

Saltz, E., & Newman, S. E. The effect of prior learning of symbols on performance in reasoning. Amer. J. Psychol., 1960, 73, 91-99.--502.

Schulz, R. W., see Underwood, Runquist, & Schulz.

Sengbush, Gertrude A. An experiment with the use of lantern slides in the teaching of typewriting. Master's thesis, University of Southern California, 1933. U--177.

Shaw, J. C., see Newell, Shaw, & Simon.

Sheffield, F. D. Theory and implications relevant to the teaching of complex response sequences by demonstration and practice. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-56-14, March 1956. U--mimeo. (superceded by Sheffield, 1957b and by Ch. 2 herein)--13.

Sheffield, F. D. Perceptual mediation in the learning of organizable sequences: theory and experiment. AFPTRC, Maintenance Laboratory, Tech. Memo. ML-TM-57-14, September 1957(a). (superceded by Chs. 2 and 8 herein)--13, 107, 110, 111.

Sheffield, F. D. Theory and implications relevant to the teaching of complex response sequences by demonstration and practice. AFPTRC, Maintenance Laboratory, Tech. Memo. ML-TM-57-18, November 1957(b). (see Ch. 2 herein)--13.

Sheffield, F. D. Theoretical considerations in the learning of complex sequential tasks from demonstration and practice. 1961: Ch. 2 herein.--4, 7, 11, 471, 483, 484, 485, 491, 493, 497.

Sheffield, F. D., see also Hovland, Lumsdaine, & Sheffield; Maccoby & Sheffield; Margolius & Sheffield; Margolius, Sheffield, & Maccoby; Margolius, Sheffield, & Robinson; Weiss, Sheffield, & DuBois; Weiss, Sheffield, & Maccoby; Wulff, Sheffield, & Kraeling.

Sheffield, F. D., & Maccoby, N. Summary and interpretation of research on organizational principles in constructing filmed demonstrations. 1961: Ch. 9 herein.--157, 493, 500.

Sheffield, F. D., Margolius, G. J., & Hoehn, A. J. Experiments on perceptual mediation in the learning of organizable sequences. 1961: Ch. 8 herein.--483, 484, 485.

Shettel, H. H., & Lindley, R. H. An experimental comparison of two types of self-instructional programs for a SAGE system paired-associate task. Pittsburgh, Pa.: American Institute for Research (AFCCDD-TN-61-20), March 1961.

Shillestad, Isabel, see Sidowski, Kopstein, & Shillestad.

Shipman, Virginia, see Gropper, Lumsdaine, & Shipman.

Sidowski, J., Kopstein, F. F., & Shillestad, Isabel J. Prompting and confirmation variables in verbal learning. Psychol. Reps., 1961, 8, 401-406.--501.

Silberman, H. F., see Coulson & Silberman.

Simon, H. A., see Newell, Shaw, & Simon.

Skinner, B. F. The science of learning and the art of teaching. Harvard Educational Rev., 1954, 24, 86-97. (In TMPL, pp. 99-113)--497.

Skinner, B. F. Final report to Human Resources Research Office. Report to Human Resources Research Office, October 1957. U (TMPL abs., pp. 683-684).

Skinner, B. F. Teaching machines. Science, 1958, 128, 969-977. (In TMPL, pp. 137-158)--386, 390.

Skinner, B. F. Why we need teaching machines. In B. F. Skinner. Cumulative record. New York: Appleton-Century-Crofts, Inc., 1961.--491.

Skinner, B. F., & Holland, J. G. The use of teaching machines in college instruction. Final Report to the Fund for the Advancement of Education, August 1958. (In TMPL, pp. 159-172)--11, 390.

Smith, P. N., Aukes, L. E., & Lumsdaine, A. A. (Eds.) Conference report. AFPTRC, Training Aids Research Laboratory, Tech. Memo. TARL TM-54-1, 1954.--527.

Spence, K. W. The nature of discrimination learning in animals. Psychol. Rev., 1936, 43, 427-449.--445.

Spence, K. W., see also Taylor & Spence.

Spitzer, M. E., see Cook & Spitzer.

Stephens, A. L. Certain special factors involved in the law of effect. Columbus, Ohio: Ohio State University Press, Abstrs. of Doctoral Dissertations, 1953, No. 64, 505-511. (In TMPL, pp. 89-93)--491.

Sternberg, J., see Miller, Levine, & Sternberg.

Stolurow, L. M. The utilization of class-descriptive cues in the learning of technical information: studies in task engineering. In G. Finch & F. Cameron (Eds.), Symposium on Air Force human engineering, personnel, and training research. Washington, D. C.: National Academy of Sciences—National Research Council, 1956. Pp. 248-266.--445.

Stolurow, L. M., see also Detambel & Stolurow; Wulff & Stolurow.

Sulzer, R. L., see also Kanner & Sulzer; Lumsdaine & Sulzer; Lumsdaine, Sulzer, & Kopstein.

Sulzer, R. L., Lumsdaine, A. A., & Kopstein, F. F. The value of using multiple examples in training film instruction. Washington, D. C.: U. S. Air Force Human Resources Research Laboratories, HRRL Report No. 25, May 1952 (see Lumsdaine, Sulzer, and Kopstein; 1961: Ch. 17 herein)--193, 241.

Taylor, D. W., see Wright & Taylor.

Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. J. exp. Psychol., 1952, 44, 61-64.--420.

Thorndike, E. L. An experimental study of rewards. Teach. Coll. Contr. Ed., 1933, 51 (580).--215, 411.

Todd, H. R., see Herman, Broussard, & Todd.

Tolman, E. C. Cognitive maps in rats and men. Psychol. Rev., 1948, 55, 189-208.--24, 484.

Twining, W. E. Mental practice and physical practice in learning a motor skill. Res. Quart. Amer. Assn. Hlth., Recreation, & Phys. Ed., 1949, 20, 432-435.--419.

Underwood, B. J., Runquist, W. N., & Schulz, R. W. Response learning in paired associate lists as a function of intralist similarity. J. exp. Psychol., 1959, 58, 70-78.

Vandell, R. A., Davis, R. A., & Clugston, H. A. The function of mental practice in the acquisition of motor skills. J. gen. Psychol., 1943, 29, 243-250.--419.

VanderMeer, A. W. The economy of time in industrial training: an experimental study of the use of sound film in the training of engine lathe operators. J. educ. Psychol., 1945, 36, 65-90.--155, 177, 181.

vanOrmer, E. B., see Hoban & vanOrmer.

Walker, E. S., see Besnard, Briggs, Mursch, & Walker; Besnard, Briggs, & Walker.

Weiss, W., & Fine, B. J. Stimulus familiarization as a factor in ideational learning. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-48, December 1955. U--mimeo: Abstr. herein.--502, 525.

Weiss, W., Sheffield, F. D., & DuBois, S. Jane. Optimum methods of combining practice with filmed demonstration in teaching complex response sequences:

II. serial learning of a geometric construction task. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-56-16, March 1956. U--mimeo. (superceded by Weiss, Sheffield, and Maccoby, 1961: Ch. 4 herein).--55.

Weiss, W., Sheffield, F. D., & Maccoby, N. Combining practice with demonstration in teaching complex sequences: serial learning of a geometric-construction task. 1961: Ch. 4 herein.--474, 480, 492.

Westfield, R. L., see Mayer & Westfield.

Wiener, Norbert. Cybernetics. New York: John Wiley and Sons, Inc., 1948.--282.

Wilcoxon, F. Some rapid approximate statistical procedures. New York: American Cyanamid Co., 1949.

Witasek, S. Uber lesen und rezitieren in ihren beziehungen zum gedachtnis. Z. Psychol., 1907, 44, 161-185, 246-282.--3.

Woodworth, R. S. Experimental psychology. New York: Henry Holt and Co., 1938.--4.

Wright, S., & Taylor, D. W. Distributed practice in verbal learning and the maturation hypothesis. J. exp. Psychol., 1949, 39, 527-531.--336, 350.

Wulff, J. J. The stimulus unit in associative learning. Doctoral dissertation, Yale University, 1954. U--142, 445, 454.

Wulff, J. J. The teaching effectiveness of a filmed mechanical assembly demonstration with supplementary nomenclature training. AFPTRC, Training Aids Research Laboratory, Lab. Note TARL-LN-55-8, 1955. U--mimeo: Abstr. herein.--502, 526.

Wulff, J. J., see also Emeson & Wulff; Kimble & Wulff.

Wulff, J. J., & Emeson, D. L. The relationship between "what is learned" and "how it's taught." 1961: ~~Ch. 30~~ herein.--333, 408, 474, 480, 482, 493, 499.

Wulff, J. J., & Kraeling, Doris G. Familiarization procedures used as adjuncts to assembly-task training with a demonstration film. 1961: Ch. 10 herein. --135, 138, 474, 477, 478, 487, 494.

Wulff, J. J., Sheffield, F. W., & Kraeling, Doris G. "Familiarization" procedures used as adjuncts to assembly task training with a demonstration film. AFPTRC, Training Aids Research Laboratory, Staff Research Memo., 1954. U--mimeo. (superceded by Wulff & Kraeling: Ch. 10 herein)--141.

Wulff, J. J., & Stolurow, L. M. The role of class descriptive cues in paired-associates learning. J. exp. Psychol., 1957, 53, 199-206. (see also Ch. 29 herein)--336, 445, 465.

Wulff, J. J., & Stolurow, L. M. The role of class-descriptive cues in paired-associate learning. 1961: Ch. 29 herein.--332, 336, 494.

Yale Motion Picture Research Project. Do motivation and participation questions increase learning? (same as Lumsdaine, May, & Hadsell, 1947)

Yamaguchi, H. G., see Gladstone, Yamaguchi, Hull, & Falsinger.

Zeaman, D. Skinner's theory of teaching machines. In E. H. Galanter (Ed.), Automatic teaching: the state of the art. New York: John Wiley and Sons, Inc., 1959. Pp. 167-176. (TMPL abs., p. 689) (Presented at AFOSR Conference on the Art and Science of the Automatic Teaching of Verbal and Symbolic Skills, University of Pennsylvania, December 1958, under title, "Comments on the automation of teaching.")--491.

Zeaman, D., see also Kaess & Zeaman.

Zuckerman, J. V. Commentary variations: level of verbalization, personal reference, and phase relations in instructional films on perceptual-motor tasks. (Pennsylvania State University Instructional Film Research Program) Port Washington, N. Y.: U. S. Naval Training Device Center, Office of Naval Research, Tech. Rept. No. SDC 269-7-4, December 1949.--155.

Zuckerman, J. V. Testing with a pre-release filmstrip as a means of predicting factual learning from a training film. Washington, D. C.: U. S. Air Force Human Resources Research Laboratories, HRRL Memo. Report No. 14, November 1951: Abstr. herein.--502, 527.

Zuckerman, J. V. Predicting film learning by pre-release testing. Audio-Visual Commun. Rev., 1954, 2 (1), 49-56.--502.

Zuckerman, J. V., & Jacobs, M. Specific effects of the film "Flight Capabilities of the F86 A" as related to points covered in the film. Washington, D. C. U. S. Air Force Human Resources Research Laboratories, HRRL Memo. Report No. 10, June 1951.--527.

* U.S. GOVERNMENT PRINTING OFFICE : 1962 O—640850