Assessment of an Evaluation by the U.S. Army of Commercial Calibration Equipment (1982)

Pages
91

Size
8.5 x 10

ISBN
030932906X

Army Calibration Study Committee; Board on Army Science and Technology; Commission on Engineering and Technical Systems; National Research Council

Assessment of an Evaluation by the U.S. Army of Commercial Calibration Equipment

A Report Prepared by the

Army Calibration Study Committee
Board on Army Science and Technology
Commission on Engineering and Technical Systems
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1982

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the Councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the Committee reponsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors, according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the federal government. The Council operates in accordance with general policies determined by the Academy under the authority of its congressional charter of 1863, which establishes the Academy as a private, nonprofit, self-governing membership corporation. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine. The National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively, under the charter of the National Academy of Sciences.

This is a report of work supported by Contract No. DAAG29-82-C-0012 between the United States Department of the Army and the National Academy of Sciences.

Available from:

Board on Army Science and Technology
Commission on Engineering and Technical Systems
National Research Council
2101 Constitution Avenue
Washington, D.C. 20418

ARMY CALIBRATION STUDY COMMITTE

Leonard L. Schneider (Chairman)
Manager, ICBM Reliability and Systems Support
TRW, Incorporated
Redondo Beach, California

Robert N. Anthony
Ross Graham Walker Professor of Management Control
Harvard Business School
Boston, Massachusetts

Laurel Auxier
Manager, Corporate Instrumentation and Measurements Standards
Beckman Instruments, Incorporated
Fullerton, California

H. McIlvaine Parsons
Manager, Human Factors Project
Human Resources Research Organization
Alexandria, Virginia

Thomas Pearson
Director
Automated Technology Associates
Indianapolis, Indiana

Arthur Plourde
President
Metron Corporation
Rancho Cucamonga, California

Gunther U. Sorger
Director, Research and Development Center
Electronic Instrumentation Division
Eaton Corporation
Sunnyvale, California

Staff:

Dennis F. Miller, Executive Director, Board on Army Science and Technology
John M. Richardson, Study Director
Helen D. Johnson, Administrative Coordinator
Cheryl A. Woodward, Administrative Secretary

CONTENTS

PREFACE

The present report is the work of the Army Calibration Study Committee of the Board on Army Science and Technology of the National Research Council. The committee was formed on March 30, 1982, to review and comment on specific aspects of an ongoing Army project to procure automated calibration equipment.

The Army project itself evolved in response to a hearing before the Senate Committee on Governmental Affairs, in November of 1981, concerning the Army's objectivity in previous similar procurements. Although the cost of these procurements is not large, the principle of objective evaluation is important.

The purpose of the committee's report is to advise the Commanding General of the U.S. Army Missile Command (MICOM) by assessing the experimental and analytical aspects of the Army evaluation. The committee neither studied nor offered any judgment on the controversy over Army procurements. It did not assess the general organization of the Army project, the procurement process within the project, nor the products procured.

This report will likely have a specialized rather than general audience--one including pertinent Army decisionmakers, the Office of the Under Secretary of the Army, current and prospective vendors of equipment similar to that tested, and those concerned with oversight of Army procurement procedures.

The committee would like to express thanks to Major General Robert L. Moore, Richard L. Hartman, and Robert O. Black, of MICOM, for their courtesy and encouragement throughout the committee's deliberations. In addition, the committee appreciates the cooperation of Larry H. Johnson and Thomas P. Tytula, also of MICOM, who worked closely with the committee, responding to its questions, and providing all needed data. Finally, the committee would like to acknowledge the Army's provision of full documentation relating to every phase of its project.

Leonard L. Schneider, Chairman
Army Calibration Study Committee

SUMMARY

This assessment originated from a decision by the U.S. Army to undertake the competitive procurement of commercially available automated meter calibration systems (AMCS). These systems comprise electronic hardware and software capable of calibrating a wide variety of electrical and electronic meters. With appropriate software, they can also instruct the operator to adjust out-of-tolerance instruments rapidly and accurately and can diagnose faults.

THE PROJECT OF THE ARMY

The Army's interest in automated calibration systems arises from its need to calibrate periodically the many and varied kinds of test, measurement, and diagnostic equipment used to maintain and repair systems essential to military operations.

Some background is necessary to appreciate the particular project that the Army undertook to reach a procurement decision for these automated calibration systems. For some time a manufacturer of these systems had alleged that the Army was slow to purchase such systems in general, and his equipment in particular. These allegations led to hearings by the Senate Committee on Governmental Affairs on November 5, 1981. At those hearings, the Under Secretary of the Army offered to have the Army make a fair, objective test of automated calibration equipment.

The project was assigned to the U.S. Army Missile Command on November 10, 1981. The Missile Command proposed a project that involved (1) requesting every known manufacturer of automated calibration equipment to supply one off-the-shelf AMCS for testing; (2) purchasing one AMCS from each manufacturer that made an acceptable response to this request; (3) making a comparative evaluation of these AMCSs; and (4) procuring for operational use larger quantities (at least seven) of one or two of the systems tested, should they prove to be advantageous relative to the existing manual equipment.

Two requests for proposal were issued to achieve these objectives. The first defined the technical requirements for an AMCS. The purchase of the AMCSs that qualified for comparative testing was based on the response to this solicitation. A second request for proposal defined the requirements for follow-on procurement of larger quantities. Among other important specifications, the second solicitation presented the evaluation criteria that were to form the basis for selection.

Three manufacturers responded to the initial request. The three AMCSs were purchased and the test was conducted.

The National Research Council was asked to advise the Army, within the framework of the Army's procurement project, whether the evaluation process was technically sound. Specifically, the Council was asked to assess the test design for evaluating the competing AMCSs, the methodology for analyzing the data generated in the comparative testing, and the Army's final report of the tests.

From the outset the Army imposed certain limitations on the procurement project. Chief among these constraints was the urgency with which high Army authority wanted the project completed. An unusually tight schedule was adopted to permit finishing the entire project within six months. The budget allotted to the actual conduct of the test also constituted a constraint. In addition, the Army desired to include a determination of cost-effectiveness. These limitations affected the scope of the project, the conduct of the test, and the evaluation process.

The scope of the project was narrowed to consideration of (1) the calibration needs of 27 fixed sites managed from the continental United States, (2) meter calibrators, (3) off-the-shelf calibration equipment, and (4) only one AMCS per vendor.

The test procedures were affected by the constraints in the following important ways:

- The test concerned only the calibration of the 15 most common types of meters in the Army's inventory.
- The experimental portion of the test was limited to determining the time needed to perform calibrations of the meters as required by the manual test procedures. Shorter calibration time was the only one of the potential advantages of the automated meter calibration equipment that could be measured with acceptable confidence within the test schedule.
- The testing was conducted at a single station.
- Surrogate operators, namely, military calibration specialists, performed the tests rather than the civilians who would actually operate the equipment at the test sites.

The evaluation methodology was limited to the use of three criteria: (1) the life-cycle cost advantage of the automated equipment, (2) the equipment performance, and (3) the results of an opinion survey of the operators testing the equipment. However, the main factor determined experimentally was the saving in time each AMCS achieved in calibrating meters, as compared with manual calibration.

Three consequences of these various constraints were the following:

- The test provided extensive data on only one of several possible advantages of the competing AMCSs--checking the calibration of a meter. The values and weights given to other functions, such as adjustment and diagnostics, had to be estimated from limited data, thus detracting from the project design's original intent of maximal objectivity.
- The ability to generalize from the test results was diminished. In particular, consideration could not be given to the abilities of these AMCSs to calibrate types of instruments other than meters.
- The differences between the test and the field situation were great enough to warrant some caution as to the validity of the results.

THE ASSESSMENT BY THE COMMITTEE

As charged by the Army, the committee reviewed the experimental design and procedures for the tests the Army developed to evaluate AMCSs. This review considered attributes and variables important to equipment performance, factors that might bias the results if uncontrolled, and the adequacy of experimental control by randomization and other methods. The committee also reviewed the quantities to be measured by the experimental design and their importance in evaluating the AMCSs.

Further, the Army's methodology was reviewed, focusing on the evaluation criteria and their assigned weights. The committee examined the soundness of the methodology as a basis for arriving at a procurement decision. It also considered factors that were not embraced by the methodology but that might bear on the procurement decision.

Last, the committee assessed the Army's final report of its tests and evaluation of the AMCSs.

The following overall, summarizing conclusion addresses the significance of the evaluation for both the present procurement and the more general future uses mentioned by the Under Secretary of the Army in his Senate testimony. (See Chapter 1.)

The project taken as a whole developed considerable information pertinent to a decision in the present procurement. In particular, the results of the formal evaluation model fairly display gross differences among automated meter calibration systems as an aid to a decisionmaker. If these results are supplemented by systematic scoring of potential, but unmeasured, capabilities of each automated meter calibration system, the Army can very likely make the procurement decision that best serves its requirements for meter calibration and the associated functions that occur in actual operations. However, this particular project suffered from too many constraints to serve as a good prototype for more general procurement evaluations of automated calibration systems.

Five specific conclusions about the design and conduct of the test and about the evaluation methodology support the overall conclusion. These specific conclusions are stated below, each followed by a short discussion.

The design of the experiment to observe the times for the procedural steps necessary to calibrate meters by the manual and the automated meter calibration systems was sound as judged by established scientific principles, and the experiment was conducted according to carefully prescribed procedures; therefore, the experimental results can be relied on to yield dependable information, subject to uncertainties associated with modifications to the data introduced to achieve commensurability of the test observations.

The times taken to calibrate meters were important inputs to the cost advantage term in the vendor evaluation model. Whereas other inputs to the cost term had fixed values, it was essential to determine the calibration times experimentally to a reasonable level of confidence. Under constraints on time and funding, the experiment chosen was a modified factorial design, with limited replication. Such a design is quite capable of reliable results when the size of the experiment is as large as it was here. A number of variables inherent to the test situation were adequately controlled through counterbalancing and randomization. Some caution is warranted as to the degree that the experimental situation is assumed to represent the field situation. (See Chapters 3, 4, and 5.)

Apart from the calibration times, information that was gathered during the test for use in the evaluation methodology--such as that on the reliabilities and calibration stabilities of the automated meter calibration systems, that on their abilities to perform tasks beyond the calibrations measured in the experiment, and that from the survey of operator opinion--had limited validity; although corresponding factors were assigned relatively low weights in the evaluation methodology, the uncertainties associated with them could impair discrimination between two systems with closely spaced scores.

Time and funding constraints limited the effort to determine the AMCS capabilities used in the vendor model other than the time it takes to calibrate meters. The information on these capabilities was derived from a very small number of observations of a large universe. Inferences drawn from small samples, of course, have a high risk of not representing well the true values of the population. Individual weights assigned to these capabilities in the evaluation model were low, ranging from 4 percent to 14 percent. Nevertheless, the resulting uncertainties in combination could blur the overall evaluation scores. (See Chapters 3, 4, and 5.)

The evaluation model did not include all information pertinent to the merit of automated meter calibration systems because such information could not have been properly quantified for a rigorous, formal model given the constraints of time and funding imposed on the project; therefore, consideration should be given to developing a score for other qualitative factors in order to optimize the procurement decision.

There are numerous characteristics of an AMCS that can be evaluated. However, the only statistically significant data that could be collected on these in the present study were the times the AMCSs take to calibrate meters. Although some additional information on several other aspects of the performance of the AMCSs was obtained, many of their potential advantages were not studied at all because of the constraints imposed on the project. The economic value of these other features was not evaluated, but it could vary widely from one AMCS to another. Even when their benefits are considered in relation to their costs, the extent of the unknown variations in economic and operational utility could perturb the rankings obtained from the formal evaluation model. Omission or inadequate weighting of some of the performance advantages would not change the ranking if the AMCS with the highest score were also best with respect to these performance factors. If, however, the vendor ranked second or third offered a system with superior performance for these features, proper consideration of them might move one of these vendors into first place. Thus, the decisionmaker should take these factors into account and should obtain additional information, perhaps by the use of panels of impartial experts making independent judgments according to explicit criteria. (See Chapter 2.)

The report of the Army on the cost-effectiveness analysis of fixed-site automated meter calibration systems was generally acceptable, appropriately describing the test plan, the evaluation method, and the evaluation results; the exposition was impaired, however, by lack of a description of the modifications to the calibration time data and the consequences of such modifications.

1

BACKGROUND

Electrical and electronic meters constitute about one-fifth of all test, measurement, and diagnostic equipment (TMDE) used by the U.S. Department of the Army. These meters, as well as other categories of TMDE, require periodic calibrations to verify their suitability for service in maintaining and repairing various systems essential to military operations. The use of automation in calibrating TMDE could offer significant advantages in the form of increased productivity and technical improvements in calibration and associated functions. To estimate these potential advantages, the Army would need to evaluate automated calibration systems in comparison with the manual systems in current use.

To appreciate the recent study that was actually undertaken by the Army and that led to this report, some additional background information is necessary. For some time, a manufacturer of automated calibration systems had alleged that the Army was slow to purchase such systems in general, and his equipment in particular.* The wide publicity his criticisms received ultimately led to hearings before the Senate Committee on Governmental Affairs on November 5, 1981. At those hearings, the Under Secretary of the Army committed the Army to conduct a "fair objective test" of automated calibration systems.** The outcome of such a test was to be "a fair set of specifications or measures and an evaluation process ... that I could count on to give me an objective impartial representation." The Under Secretary

*Letter from Milton J. Socolar, Acting Comptroller General of the United States, to the Honorable Joseph P. Addabbo, Chairman, Subcommittee on Defense, Committee on Appropriations, U.S. House of Representatives, April 3, 1981.

**U.S. Congress, Senate, October 21, 27, and November 5, 1981. Acqusition Process in the Department of Defense, Hearings before the Committee on Governmental Affairs. Washington, D.C.: U.S. Government Printing Office, page 609.

stated an additional desired outcome, namely, that he "would like also to make this a test of something that would wind up going on into future use so that we could get more than just that one test benefit of it." As a direct consequence, immediate attention was given to designing and carrying out a test to evaluate automated calibration equipment that would satisfy the Army's requirements.

The National Research Council (NRC) was asked to assist with part of the evaluation process to assure that it was technically sound. Specifically, the NRC was to examine the design for testing the competing equipment and the methodology for analyzing the data generated in the comparative testing.

THE PROJECT FOR PROCUREMENT OF AUTOMATED METER CALIBRATION SYSTEMS

The test, measurement, and diagnostic equipment of the Army includes not only direct-current (dc) and alternating-current (ac) meters, but other electronic equipment such as oscilloscopes, signal generators, spectrum analyzers, frequency counters, microwave measurement instruments, mechanical gauges, and radiation survey meters. However, in this procurement the Army chose to focus solely on automated meter calibration systems (AMCS) for dc meters and ac meters operating up to 10 MHz. The stated reasons were that there appeared to be sufficient workload relating to these meters for cost-effective automation, that such automated systems were commercially available, and that the existing inventory of manual calibration equipment would benefit by the addition of a capability to calibrate ac ammeters, a capability AMCSs can provide.

The task was assigned to the U.S. Army Missile Command (MICOM) on November 10, 1981; its details were established in cooperation with higher authority by December 3, 1981. A project management plan was adopted with the following stated purpose:

> The purpose for this program is to evaluate off-the-shelf commercial automatic meter calibration equipment and select, for acquisition, the most cost effective system to satisfy the Army's immediate requirement. This will be accomplished by purchasing candidate contractor calibration equipment, conducting comparative testing and analytically determining the most cost effective approach for augmenting existing manual instrumentation for this generic class of TMDE.
>
> -- Appendix C, Item 2.

The same management plan sets forth the following objectives:

1. Identify Army requirements for automatic meter calibrations and prepare a RFP [Request for Proposal] for industry proposed solutions.

2. Purchase one set of each acceptable industry proposed calibration system for evaluation.
3. Conduct comparative tests to obtain automatic meter calibration system performance data and determine the extent to which all systems meet the contract requirements.
4. Purchase and demonstrate one set of maintenance software for the performance testing, adjustment and basic repair functions for one representative TMDE test unit.
5. Conduct an appropriate analysis to determine the most cost effective automatic meter calibrator for augmenting existing manual instrumentation for this general class of TMDE.
6. Procure the most cost effective equipment.

-- Appendix C, Item 2.

The objectives of the project management plan were to be achieved through a sequence of steps. An initial solicitation (Appendix C, Item 3), hereafter referred to as Request for Proposal (RFP) No. 1, defined the technical requirements for the fixed-site AMCS; its software; and the training, maintenance, and repair required to support the system. The Army would procure one unit of each AMCS qualifying under RFP No. 1. These AMCSs would be tested in operation as meter calibrators to yield data such as the times needed to calibrate meters, the frequency of required recalibration of the AMCSs themselves, and the dependability of their operation. The AMCSs would then be evaluated with respect to life-cycle cost advantage, equipment performance, and the opinions of the operators about them. Of the 25 firms that received RFP No. 1, only three proposed to supply AMCSs. All three were considered responsive to the RFP, and the Army subsequently purchased one AMCS from each of the three vendors.

In the meantime, a second request for proposal, RFP No. 2 (Appendix C, Item 4), was issued that defined the requirements for follow-on procurement of larger quantities of at least seven. Only the three companies that supplied AMCSs in response to RFP No. 1 were allowed to bid on RFP No. 2. In addition to restating the technical requirements for the fixed-site AMCS, RFP No. 2 prescribed the quantities of AMCSs that were to be procured at the option of the Army, the evaluation criteria that were to form the basis for their selection, the training to be provided for operators of the equipment, the maintenance to be provided, and the delivery schedules. The project would then be completed with the decision to award, or not to award, on the basis of the evaluation and the response to RFP No. 2. To give some idea of the size of the procurement, the initial equipments were purchased at unit prices ranging from about $20,000 to about $70,000.

THE TASK OF THE COMMITTEE

The committee was established at the request of the Commanding General, U.S. Army Materiel Development and Readiness Command, and with the concurrence of the Under Secretary of the Army.

Within the framework of the established procurement plan outlined above, the task of the committee was to assess the following:

- o The experimental design and test procedures used by the Army in testing commercial calibration equipment.
- o The methodologies used in the analysis of the data.
- o The Army's final report of its tests.

The purpose of this task was to provide an independent assessment of the technical soundness of the comparative evaluation and selection of AMCSs. The principal features of this process had been formulated, adopted, and scheduled prior to the committee's assessment. The full statement of the task is given in Appendix A.

The manufacturer who had raised objections to the procurements of the Army did not respond to RFP No. 1. Consequently, his equipment was not included in the test, and the test did not evaluate that equipment. Upon an invitation by the National Research Council to provide comments relevant to the task of the committee concerning the test procedure, the manufacturer provided an explanation of his decision not to respond together with other material. The committee read this material, and considers that those points in it relevant to the committee's task are adequately covered in this report. In addition, the committee wants an important point to be clear: the disputes of individual vendors are outside the scope of its charge.

THE ACTIVITIES OF THE COMMITTEE

The committee consisted of individuals with expertise in the fields of calibration, engineering design of calibration equipment and systems, test design, statistical analysis, analysis of benefit in relation to cost, software life-cycle cost, human factors analysis, reliability, and quality assurance.

A list of the meetings of the committee is given in Appendix B. An initial two-day meeting was held at Redstone Arsenal, Alabama, the site of the equipment tests. At this meeting, the committee was briefed by the Army on the background and implementation of the procurement and on the procedures for the comparative test and evaluation of the AMCSs. The committee visited the test site and observed technical personnel operating the three makes of AMCS purchased under RFP No. 1, as well as the manual calibration equipment in current use.

After studying the extensive written materials provided by the Army, the committee met a second time for discussion. Further oral and written material was provided by the Army. Two subsequent meetings served to develop and refine drafts of the committee's report. Individual committee members conferred extensively with Army personnel on matters of factual accuracy and clarity.

The Army cooperated fully with the committee by providing extensive documentation on the procurement project. This material covered relevant background information, the project management plan, and the two RFPs. The material also covered the specific plan for conducting the experimental part of the test, including some actual test data and results. The material described the methodology for evaluating the overall merit of the AMCSs, specifying various inputs to the evaluation such as calibration workloads at the 27 designated calibration sites in the continental United States, applicable labor rates, discount factors for money, and the text of the operator questionnaires that were used to ascertain operator opinion. In addition, the vendors' specifications for their hardware were provided to the committee. Relevant portions of the final report of the Department of the Army Test, Measurement and Diagnostic Equipment Action Team (DATAT)--the so-called "Bonner Report"--were made available. Lastly, the Army's final report of its tests was made available for the assessment of the committee. These materials are identified in full in Appendix C. Where the content of a particular document is crucial to the findings and conclusions of the committee, the pertinent portion of the document is reproduced as an appendix for easy reference.

2

CONSTRAINTS IMPOSED ON THE ARMY PROJECT

The committee noted that a number of constraints imposed on the Army's procurement project affected procedures and decisions relating to the project's scope, to the conduct of the test, and to the evaluation methodology.

CONSTRAINTS ON THE SCOPE OF THE PROJECT

The time for accomplishing the project was shortened from the originally recommended 14 months to 6 months. This exceptionally short period the project was allowed resulted from the sense of urgency conveyed by the Headquarters of the U.S. Army Materiel and Readiness Command (DARCOM). A budget of $1.35 million was made available by DARCOM for the project, $300,000 of which was allocated to the actual conduct of the test. The project was further restricted to consideration of domestic sites because an Army committee (the "Bonner Committee") was studying the optimal way of performing the calibration function at mobile and overseas locations (Appendix C, Item 7.) Specifications for the equipment at these locations could not be developed until this study was completed.

These constraints narrowed the scope of the project in four ways, limiting the analysis to (1) the calibration needs of 27 fixed sites managed from the continental United States; (2) the calibration of meters; (3) off-the-shelf calibration equipment, to meet the 47-day delivery time specified in Request for Proposal (RFP) No. 1; and (4) only one automated meter calibration system (AMCS) per manufacturer. It is not possible to judge how seriously these restrictions impair the applicability of the results to field operations.

CONSTRAINTS ON THE TEST PLAN

Constraints on the Instruments Tested

Under the time and budgetary constraints, a decision was made to limit the test to calibration of meters. One justification for this limitation was that meters represent the largest fraction of electronic test devices used by the Army. Additionally, the inclusion of other electronic test equipment, such as oscilloscopes and microwave instruments, would unduly lengthen the time to obtain and analyze the AMCSs because of the greater complexity of systems capable of calibrating such devices.

In compliance with DARCOM's directive that a representative sample of the Army's population of test, measurement, and diagnostic equipment (TMDE) be selected for testing, the test was limited to the 15 meter types most common in the Army's universe of 354 types. The selected meters represented 80 percent of the total number in the Army's inventory, but not necessarily the same proportion of the calibration workload.

All automated calibration equipment requires the development of software tailored to the calibration of a particular instrument. Thus, the limitation to 15 specified meter models was also necessary so that the software requirements specified in RFP No. 1 could be met within the short delivery time available.

Constraints on the Test Observations

One vendor's AMCS may differ from that of another in at least the following respects, all of which are in principle susceptible to observation:

1. The time it takes to calibrate meters.
2. The time it takes to adjust meters that are out of tolerance.
3. The usefulness of its diagnosis of a defective meter (and the effect of this information on the time required for repair).
4. The uncertainties of its calibration, adjustment, and diagnosis operations (a function of both the hardware and software of the AMCS).
5. Its ability to produce information useful for management purposes.
6. The skill it requires of operators.
7. The fatigue it produces in operators.
8. The cost of training its operators.
9. Its safety.
10. The number of meter types in the present inventory that it can calibrate, adjust, and diagnose.
11. Its ability to calibrate, adjust, and diagnose newly developed meter types, including those fitted with the Institute of Electrical and Electronics Engineering type 488 Instrumentation Bus, and the cost of developing programs for it to carry out these functions.
12. Its ability to calibrate, adjust, and diagnose test equipment other than meters.

13. Its acquisition cost.
14. Its useful life.
15. Its downtime resulting from maintenance, repair, and other causes.
16. Its reliability, that is, its mean-time-between-failures.
17. The frequency with which the AMCS itself must be recalibrated, together with ease and speed of recalibration.
18. Its costs for maintenance and repair.

The only information that could be collected at a satisfactory level of statistical confidence in the time available was that on the time a system required to detect whether a meter was in or out of tolerance--that is, the time a system required to check meter calibration.* Some additional information was obtained on other aspects of the performance of each AMCS, but it is not statistically significant. Also, the costs for acquisition, maintenance, and training for each AMCS were obtained directly from the proposal received in response to RFP No. 2.

Additional Test Constraints

The need for rapid test results, the budgetary limitations, and DARCOM directives all acted to impose further less-than-optimal test conditions: (1) the test was conducted at only one station; (2) adjustment and repair capabilities were limited to a single-sample demonstration; (3) vendors were afforded only a limited opportunity to demonstrate additional AMCS capabilities; (4) in accordance with DARCOM's directive, the test operators were enlisted calibration specialists rather than the civilian employees who would actually operate the equipment at the test sites; and (5) there was no means to test whether the single AMCS from each vendor was truly representative of all units of the same model.

CONSTRAINTS ON THE EVALUATION METHODOLOGY

As noted in Chapter 1, the Under Secretary had committed the Army to conducting a fair, objective test. Consequently, subjective factors were minimized in the scoring. Efforts to put the evaluation on a numerical basis had the effect of limiting the observed characteristics to easily measurable quantities and of excluding from consideration a number of important but less easily quantifiable characteristics. Another consideration that influenced the nature of the evaluation methodology was a desire to determine cost-effectiveness.

With these points in mind, the evaluation criteria for selecting the equipment to be procured were established, and set forth in RFP No. 2.

---

*The term calibration may refer generally to a variety of functions necessary to ensure the accuracy of a measuring instrument. However, with respect to the observations associated with the Army's test, the term refers to whether a meter is in or out of tolerance.

Specifically, three factors were to be evaluated: (1) life-cycle cost advantage, (2) equipment performance, and (3) a survey of the operators' opinions of the AMCSs. Once announced in RFP No. 2, these factors themselves became constraints on the evaluation methodology.

In measuring the cost advantage, the cost of making calibrations with existing manual equipment was used as a benchmark. For each competing AMCS, the time required to calibrate meters was measured; and its saving in time over manual calibration was determined. This saving was transformed into a monetary amount of savings over 4.5 years, a consequence of a 5-year maximum permissible term for TMDE maintenance contracts. The costs of purchasing and maintaining the equipment and of training operators were then deducted from the present value of this stream of savings. The AMCS with the greatest cost advantage (or the least cost disadvantage) received the highest score.

RFP No. 2 specified the performance characteristics to be evaluated as frequency of recalibration, dependability, potential ability to calibrate additional meters, and potential diagnostic capability.

The opinion survey was to be used to estimate the ease and adequacy with which personnel could be trained and could routinely operate the AMCSs.

The principal constraint imposed by RFP No. 2 was in how these three evaluation factors were to be weighted in arriving at the overall cost-effectiveness of each AMCS. Cost advantage was to receive the highest scoring weight, and equipment performance a higher weight than the opinion survey.

SOME CONSEQUENCES OF THE CONSTRAINTS

The various constraints on the project, and thus on the test, had several significant implications. The emphasis on fairness and objectivity led directly to efforts to rely on quantitative results. The test provided objective data and, consequently, objective scoring on several of the possible advantages of AMCSs. However, the weights given to the various scores entering into the evaluation methodology were arrived at judgmentally. Thus, although the project was intended to be objective, some judgment did enter the evaluation of the test results.

As a consequence of the desire to complete the project within six months, an unusually tight schedule was adopted. The RFP to acquire AMCSs was issued on January 14, 1982; the three vendors that submitted proposals were awarded contracts calling for the delivery of their AMCSs on March 29, 1982; evaluation of the AMCSs was to be complete by early June 1982; and a contract for quantity procurement was to be awarded on June 14, 1982 (later changed to October 1982).

There were three noteworthy consequences of the tight schedule and resulting necessity to impose certain limitations on the test: (1) the test provided extensive data on only one of several possible advantages of the competing AMCSs, namely, checking the calibration of a meter; (2) the ability to generalize from the test results was limited; and (3) the differences between the experimental and the field situation are great enough to warrant some caution as to the validity of the results. Specifically, limiting the analysis to meter calibration prevented consideration of the ability of the AMCSs to calibrate other types of instruments; and limiting the analysis to off-the-shelf AMCSs precluded the possibility of evaluating new equipment tailored to the Army's requirements.

3

ASSESSMENT OF THE TEST PLAN

The experimental design and test procedures that the Army developed to evaluate the performance of automated meter calibration systems (AMCS) are set forth in Item 5 of Appendix C. The committee reviewed the design and procedures with attention to the attributes and variables important to equipment performance, the factors that might bias the results if uncontrolled, and the adequacy of experimental control by randomization and other methods.

THE ITEMS TESTED

The item being tested is the most important element in any test of equipment. In this case, the tests were performed on commercially available AMCSs that the responding vendors felt best met the requirements of Request for Proposal (RFP) No. 1. (See Appendix C, Item 3.)

The first specification in this RFP was the type of AMCS. The Army specified that the AMCS be capable of calibrating general purpose meters in the Army's inventory. Other general specifications were that the AMCS should be capable of operating in a manual mode, and that its method of calibration should be similar to currently used methods documented in Army Technical Bulletins for these meters.

If a vendor believed his equipment could perform at least some, if not all, of the calibration steps required for a set of representative meters, he was still qualified to respond provided he listed those steps his AMCS was incapable of performing.

The AMCS was also to possess additional capabilities for further applications, and the control of test, measurement, and diagnostic equipment (TMDE). However, the AMCS was not to be so elaborate or costly as to put it beyond budgetary restrictions. In addition to calibrating the meters, the equipment was required to support adjustments and diagnostics on faulty meters.

The AMCS had to operate in a specific manner so as to return to safe states at logical entry and exit points, to evaluate and display errors in the meter it calibrates, and to provide options for the operator should the meter fail to meet the tolerance limits of a specific calibration step. Finally, certain calibration and reliability requirements were imposed.

To further assist the vendor in making a choice of which of his AMCSs he should propose, the Army included an outline of its tentative test and evaluation plan for the procurement as Attachment D to the RFP. The test plan indicated there would be two major sections of the test. The first section specified characteristics of a general nature, and consisted of three major parts:

1. A measure of the performance of the AMCS against standards of electrical quantities. This part consisted of measuring the accuracy of the values of voltage, current, and resistance used to excite the meters. (According to Army records of this section of the test, for dc voltage and current, the AMCSs were accurate to a few tens of parts per million with respect to the national standards. For dc voltage, ac current, and resistance, the AMCS accuracies were at least 4 times the rated accuracy of the meters under test. The various meters being calibrated had rated accuracies in the range of 0.1 percent to 10 percent.)
2. Verification of vendor claims. This part provided that the representations in the literature supplied by the vendor describing the capabilities of his equipment would be verified.
3. A determination of the degree to which the AMCS met the contract requirements as specified in Attachment A to the RFP. Attachment A specifies the 21 following requirements:

    1. Function of AMCS.
    2. Automatic/manual operation.
    3. Interface/display characteristics.
    4. System components.
    5. Environmental range within which specified operation shall occur.
    6. Military Standard 454.
    7. Electrical power requirements.
    8. Accuracy and reliability of calibration.
    9. Software medium other than paper tape.
    10. IEEE 488 Instrumentation Bus.
    11. Calibration procedure for AMCS.
    12. Ability to calibrate worldwide Army meter population.
    13. Accuracy 4:1 with respect to the 15 meters in the test.
    14. Explanation of method of calibration by AMCS.
    15. AMCS to follow method of Army Technical Bulletins for meter calibration.
    16. Safe condition at entry and exit points.

17. Display of operator instructions.
18. Control of stimulus by AMCS.
19. Pass/fail indications.
20. To proceed if meter passes calibration.
21. To halt if meter fails calibration and to provide decisions to operator: repeat, continue, or terminate.

The Army also conducted a human engineering assessment of the three AMCSs. The results were reported in an appendix to the Army's final report, but they were not incorporated into the scoring model.

The three AMCSs tested varied widely in design and capabilities within the rather broad specifications set forth by RFP No. 1. Because of this fact, it would have been difficult to evaluate them in all respects by absolutely uniform criteria.

The purpose of the second section of the test was to provide the information on calibration times and seven other aspects of the performance test, as discussed in the section below.

THE TEST OBSERVATIONS

The test observations described below provided good quantitative data on the procedural step times for calibration, but weaker quantitative and qualitative data on seven other characteristics of the AMCSs.

Experimental Design of the Observations of Times Taken for Calibration

A fairly elaborate experiment was conducted to observe the procedural step times for calibration. The experimental design, a somewhat-modified factorial, is presented in two different ways in Tables 1 and 2 to facilitate comprehension. The tables are derived from Army documentation and are presented to give a clearer exposition of the design. The three AMCSs constituted the principal independent variable of interest to the Army. Five other variables in the formal design were nine operators, fifteen meter models, three AMCS cycles (for sequencing purposes), six sets of meters, and two tolerance conditions (included among the six meter sets). These five variables were systematically manipulated for purposes of experimental control. The fifteen meter models were common to each of the six sets of meters; that is, there were different meters in the sets, but the fifteen meter models in each set were the same. In three of the sets, most of the meters included were out of tolerance. The out-of-tolerance conditions were all distinct enough to be unambiguously detected by the AMCSs. Hence discrimination among AMCSs was on the basis of the time, rather than the ability, to detect an out-of-tolerance condition.

TABLE 1 Structure of the Test Design

System Factors	Cycle 1	Cycle 2	Cycle 3	Cycle 4
Manual System 1				
Operators	1, 5, 6			
Meter Sets	A, E[a]			
Manual System 2				
Operators	2, 3, 7			
Meter Sets	B, D			
Manual System 3				
Operators	4, 8, 9			
Meter Sets	C, F			
Automated Meter Calibration System 1				
Operators		1, 5, 6	4, 7, 8	2, 3, 9
Meter Sets		B, D	C, E	A, F
Automated Meter Calibration System 2				
Operators		3, 7, 9	1, 2, 6	4, 5, 8
Meter Sets		C, F	A, D	B, E
Automated Meter Calibration System 3				
Operators		2, 4, 8	3, 5, 9	1, 6, 7
Meter Sets		A, E	B, F	C, D

[a]Underlined letters represent meter sets in which meters are out of tolerance.

TABLE 2 Filled Cells of the Test Design

Calibration System and Operator	Cycle 1						Cycle 2						Cycle 3						Cycle 4					
	In Tolerance			Out of Tolerance			In Tolerance			Out of Tolerance			In Tolerance			Out of Tolerance			In Tolerance			Out of Tolerance		
	A[a]	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
Manual System																								
Operator 1	X				X																			
Operator 2		X		X																				
Operator 3		X		X																				
Operator 4			X			X																		
Operator 5	X				X																			
Operator 6	X				X																			
Operator 7		X		X																				
Operator 8			X			X																		
Operator 9			X			X																		
Automated Meter Calibration System 1																								
Operator 1								X		X														
Operator 2																			X					X
Operator 3																			X					X
Operator 4															X		X							
Operator 5								X		X														
Operator 6								X		X														
Operator 7															X		X							
Operator 8															X		X							
Operator 9																			X					X

[a] Letters A through F indicate the six meter sets.

TABLE 2 Filled Cells of the Test Design (Continued)

Calibration System and Operator	Cycle 1						Cycle 2						Cycle 3						Cycle 4					
	In Tolerance			Out of Tolerance			In Tolerance			Out of Tolerance			In Tolerance			Out of Tolerance			In Tolerance			Out of Tolerance		
	A[a]	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
Automated Meter Calibration System 2																								
Operator 1													X			X								
Operator 2													X			X								
Operator 3									X			X												
Operator 4																				X			X	
Operator 5																				X			X	
Operator 6													X			X								
Operator 7									X			X												
Operator 8																				X			X	
Operator 9									X			X												
Automated Meter Calibration System 3																								
Operator 1																					X	X		
Operator 2							X				X													
Operator 3														X				X						
Operator 4							X				X													
Operator 5														X				X						
Operator 6																					X	X		
Operator 7																					X	X		
Operator 8							X				X													
Operator 9														X				X						

NOTE: The sixth independent variable (in addition to meter system, operator, test cycle, tolerance status, and meter set) represents the 15 meter models and is common to all conditions and occupies each filled cell.

[a] Letters A through F indicate the six meter sets.

Table 2 shows 54 cells comprising combinations of the variations in AMCSs, subjects, cycles, and meter sets (including the tolerance variable). When the 54 cells in Table 2 are multiplied by the 15 meter models in each cell, it can be seen that the entire design consisted of 810 cells. Cells are empty either because combinations were not repeated across cycles or because operators encountered different (but equivalent) meter sets with different AMCSs (to prevent memorization of meter calibration status). These two modifications of a full factorial design (requiring nine times as many cells) do not importantly affect the experimental results. They were also advisable in view of the constraints on the project, especially given that more than three vendors could have responded.

The operators were selected so to have similar pertinent skills. The electromechanical features of meters of the same model were also presumed to be similar. Given the limitations of time and funding, this test design permitted the data necessary for the evaluation to be collected.

The principal measure obtained was the time taken to make a group of calibrations, which were summed for all the groups of calibrations carried out on a given meter. (See Table 3, also presented to clarify material in Army documents.) For each set of 15 meters, there were 573 inputs (and values read), with a range of 22 to 60 per meter model. The AMCS comparisons (not including the manual system) represented 30,672 inputs and resultant readings of values (including 1,620 out of tolerance). In each of three sets (D, E, and F), there were 20 out-of-tolerance faults, with a range of 0 to 15 per meter model. With the three exceptions noted in Table 3, the specific faults differed from set to set.

For each set of 15 meters, there were 56 time data points, with a range of 2 to 5 per meter model. The AMCS comparisons produced 3,024 time data points. Because the data points were summed for each meter, 810 rather than 3,024 data points were used in the data analysis.

The Army recognized that all vendors would not have an AMCS capable of performing each and every task on all the representative meters. Therefore, it would be very difficult to evaluate relative performance. To correct this deficiency, the measured times for automated calibration were adjusted in two ways. If it was clear that an AMCS could perform the missing tasks given simple changes in its software, then a correction was calculated from the observed automated calibration times. For example, if 3 points on a scale were actually checked and 5 points on the same scale were required to be checked, then the adjusted time for the required 5 points was taken to be 5/3 of the observed time. If an AMCS could not perform one or more tasks on a given meter, the time to calibrate that meter was measured by the sum of the times the AMCS required to do those tasks that it could plus the time required by the manual system to do those tasks that the AMCS could not.

The experiment yielded both performance times for all systems, and differences in time between the manual system and each AMCS. From

TABLE 3 Meter Model Variable

Meter Model	Inputs (Values Read) F1[a]	F2	F3	F4	F5	Total	Data Points	Faults D[b]	E	F	Total
1	18	10	3	8	5	44	5	1[c]	1[c]	1[c]	3
2	10	9	5	-	-	24	3	-	-	-	0
3	19	7	8	5	-	39	4	1	1[d]	2[d]	4
4	11	6	4	7	-	28	4	2	2	3	7
5	16	48	-	-	-	64	2	2	2	2	6
6	6	15	5	6	8	40	5	-	-	-	0
7	8	6	10	3	-	27	4	2	1	-	3
8	16	24	-	-	-	40	2	2	2	-	4
9	15	7	8	11	8	49	5	4	3	8	15
10	16	14	-	-	-	30	2	2[d]	4[d]	1	7
11	12	7	6	7	-	32	4	2	-	-	2
12	5	10	5	6	15	41	5	1	3	1	5
13	10	4	4	4	-	22	4	-	-	-	0
14	13	47	-	-	-	60	2	1	1	2	4
15	13	6	2	6	6	33	5	-	-	-	0
Total						573	56	20	20	20	60

[a]The five functions for which meters were tested were (1) dc volts, (2) ac volts, (3) frequency response, (4) dc current, and (5) resistance.

[b]D, E, and F represent the three meter sets that included meters out of tolerance.

[c]The fault in all meters was the same.

[d]One fault is the same for the two meters.

either measure, it was possible to derive differences among AMCSs, among cycles (to indicate any learning or fatigue effects), among operators, among meter models (different models had different calibration needs), among meter sets (in changes within pairs of meter sets, learning might occur from the first set to the second), and between tolerance conditions, as well as interactions among AMCSs, operators, and meter models.

Internal Validity of the Experimental Design

Effects of learning, fatigue, electromechanical variations, and individual differences among operators seem to have been well controlled through the particular factorial design used and through counterbalancing and randomization.

Consideration must be given, however, to potential confounding from extraneous variables. One such variable could be the inadvertent disclosure of information about their performance times to the experimental subjects. It is known that information feedback of this sort can lead to a modification of subjects' performance (which has been construed as a form of the Hawthorne effect). In turn, the modified performance might be different for different AMCSs. To avoid this effect, the Army took precautions to prevent such feedback prior to the first session in which data comparing the AMCSs were obtained. Another such variable is the extent and quality of training received on each AMCS. These were vendor-determined, and no attempt was made to check whether operators had reached their best performance before the experiment began. (Previous operations on meters using the manual system had presumably brought operators to their best performance for that set of operations.) Further, the workplace arrangements, such as seating, could differ among systems; it was reported that one change was made to equalize these arrangements. None of these extraneous variables was likely to have had major differential results, in the committee's opinion.

Overall, in the committee's judgment, the design of the experiment enjoys commendable internal validity.

External Validity of the Experimental Design

A remaining issue is whether the test adequately represented the external context to which it was addressed. Among the components of the test program that might be examined for external validity were the program's location, the particular operators it employed, the particular TMDE used, additional AMCS applications, and the downtime of AMCSs. The committee recognized that the need for internal validity inevitably places some constraints on external validity, that external validity can never be complete in a experimental test program, and that any deficiences in the determination of cost savings would be worrisome not only if they affected the relative performance capabilities of the three AMCSs or the relative capabilities of the manual and automated systems, but also if they were seriously in error relative to the intended subsequent use of the AMCSs.

It was not possible to compare ambient and anthropometric aspects of the test site with those at the fixed sites, but no grounds for concern were apparent. The test operators were military personnel. No attempt has been made to determine whether the distribution of their skill levels matched those of the civilian operators at the fixed sites where the AMCS equipment would be used. Such a comparison would have to be made with the manual system. The committee was told, however, that many of the fixed-site calibrators had been military calibrators. In view of this information, their comparative skill in operating the equipment was not considered to be a problem in interpreting the test, when the relative performances of the AMCSs were being compared. However, the test times have to be realistic--not widely different from that occurring in field work. The TMDE in the test consisted of 15 meter models that were most common in the Army inventory. However, the evaluation extrapolated to 57 other meter models. Consideration of meter models that might replace some among the 15, including future automatic meters, was judged beyond the scope of analysis of external validity, as was the consideration of other TMDE such as oscilloscopes. Finally, in the field situation, calibration personnel may well use the AMCSs for additional applications, such as in adjustment, diagnostics, and management of the calibration process. Downtime in the field will surely differ from that in the experiment.

Observation of Other Data

Other data collected during the course of the test and immediately thereafter provided either weak quantitative or qualitative information on the following AMCS features:

1. The stability of the calibration of the AMCS itself. This stability was checked by the recalibration of the AMCS periodically throughout the testing schedule.
2. The reliability of the AMCS. Reliability was measured by a count of the number of hardware failures during the testing time of approximately 80 days.
3. The correctness of the AMCS in deciding the calibration status of a meter. Each of the 30,672 AMCS readings was checked for correctness--that is, for whether the AMCS decision that the meter was in or out of tolerance was correct. Random errors were expected to be few because the magnitude of the out-of-tolerance condition was usually large compared to the uncertainty of the meter under test. In the case of consistent errors, a scoring committee was to discard the time data on the

presumption of some meter difficulty or AMCS incapability and to substitute, respectively, either proportional AMCS measurement time or augmented time from the manual system.

4. The number of tasks the AMCS is capable of performing on the 57 untested meters in Table 4 of Attachment A of RFP No. 1.
5. The diagnostic capability of the AMCS demonstrated on one type of meter.
6. Subjective reactions to various features of the AMCSs, obtained from the opinion survey of the operators.
7. Potential additional capabilities and unique features of the AMCS for supporting Army TMDE that were neither specifically tested nor referenced in RFP No. 1.

From these seven test quantities and the procedural step time, the Army evaluated which of the AMCSs tested would be the most cost-effective for deployment at fixed sites managed from within the continental United States.

MERIT OF THE TEST PLAN

Of the eight kinds of test data described in the preceding section, the design of the experiment was adequate for determining only the relative times taken by the AMCSs in calibrating the meters. The observed times and the time differences with respect to the manual system are used as random variable inputs to a cost model. (See Appendix C, Item 6.) The cost model also includes inputs of a deterministic nature such as investment costs, wage rates, and calibration workloads.

The data from this study cover only 45 days and therefore are inadequate for determining the ability of an AMCS to maintain an in-calibration status for the desired mean time of 569 days. This point is discussed more fully in Chapter 4.

A similar statement applies to hardware failures of the AMCS, having only 330 hours of testing when the mean-time-between-failures is expected to be 536 hours. This point also is discussed more fully in Chapter 4.

The additional calibration capability of an AMCS with respect to the 57 untested meters was obtained from consideration of their calibration requirements compared to AMCS capability and thus does not rest on quantitative observations.

Diagnostic capability was inferred from a single observation, because no additional measurements could be carried out within the available test time. As such, it may be a poor measure of actual diagnostic capability.

Simple statistical methods were used to ensure that the results of the operator opinion survey were sufficiently reliable to include in the evaluation methodology. However, the validity of the opinion survey itself is questionable.

The probability of success in detecting meter errors presumably depends mainly on how far the meter is out of tolerance compared to the uncertainty of the meter setting (because the uncertainty of the AMCS is small compared to that of the meter). The conditions of the test were such as to produce a high incidence of successes. Army representatives reported that the instances of unexplained decision errors were too rare to warrant recording or processing. However, in the field closer decisions will occur; and the question of whether wrong decisions are ascribable to the meter or the AMCS is pertinent. The test was not designed to answer this question.

Qualitative indications of capabilities and features not specifically tested were clearly not subject to any statistical treatment.

To summarize, only the observations of procedural step times for calibration were amenable to statistical analysis of any appreciable power. Other observations made during the test were adequate for determining the other parameters only to a limited degree of validity because of the test constraints discussed in Chapter 2. These limitations must be recognized when the latter observations are used in the evaluation technology, as is discussed in Chapter 4.

4

ASSESSMENT OF THE EVALUATION METHODOLOGY

The methodology used by the Army for evaluating the competing automated meter calibration systems (AMCSs) is reproduced in Appendix F. This methodology is first described with respect to the evaluation criteria and their assigned weights; it is then examined for its adequacy as a decision rule for evaluating competing AMCSs for the stated purpose of procurement. In addition, some observations are made on the usefulness of the methodology for a broader evaluation of automated calibration equipment.

DESCRIPTION OF THE EVALUATION METHODOLOGY

Measures of six of the eight observed AMCS characteristics discussed in the preceding chapter were incorporated in a vendor evaluation model given by the following formula:

$$S = 0.55R + 0.40(0.35X_1 + 0.35X_2 + 0.20X_3 + 0.10X_4) + 0.05P,$$

where:

S = the vendor score.

R = $(C - C_{min})/(C_{max} - C_{min})$; C is the cost saving for a given vendor, C_{min} is the least cost saving obtained, and C_{max} is the greatest. (The greatest cost saving, namely zero, was obtained for the alternative of not procuring any AMCSs, so that $C_{max} = 0$.)

X_1 = the score for the number of recalibrations required: 0 if the AMCS was out of calibration two or more times during the test period; 0.50 if out one time; and 1.00 if it required no recalibration.

X_2 = the score for the number of times the AMCS failed during the testing period other than from causes associated with the operator: 0 for three or more failures; 0.34 for two failures; 0.75 for one failure; and 1.00 for none.

X_3 = the ratio of (a) the number of tasks that an AMCS could perform on the 57 meters listed in Table 4 of Attachment A to Request for Proposal (RFP) No. 1 that were not tested to (b) the total number of tasks required to calibrate those meters.

X_4 = in the demonstration phase of the test, for a sample of seven meters of one model only, the ratio of (a) the number of faults correctly diagnosed by the AMCS reduced by the number of faults incorrectly diagnosed to (b) the total number of known faults in that sample.

P = a normalized preference score calculated from the operator opinion survey.

According to the model, 55 percent of the weight is assigned to the cost savings evaluation of the AMCS; 14 percent to the ability of the AMCS to maintain calibration; and 14 percent to its functioning without failure. A weight of 8 percent is allowed for future or potential ability to calibrate the 57 additional meters listed in RFP No. 1, and 4 percent is allowed for correct diagnostics by the AMCS on a group of selected meters. The remaining 5 percent of the weight is allotted to the results of the opinion poll of the test operators. Details of each element of the vendor score are given in Appendix F.

ADEQUACY OF THE METHODOLOGY

In order to assess the vendor evaluation model, its elements must be examined relative to the corresponding observed AMCS characteristics noted in Chapter 3.

The Cost Model

Description of the Cost Model

The first element in the vendor evaluation model is the relative cost of the competing AMCS. According to a provision in RFP No. 2, at least seven AMCSs were to be purchased. Therefore, the cost model was designed to measure the cost savings, over manual calibrators, at the seven sites at which an AMCS would achieve the greatest savings. These sites were not necessarily the seven with the largest volume of calibration work, because the savings varied from meter to meter and the mix of meters to be calibrated was different for different sites. The model estimated savings over a 4.5-year period, because a requirement in Army procurement regulations was said to limit the maintenance contracts, and hence the firmly fixed costs, to approximately that duration.

The model used the following method to estimate the relative costs of checking meter calibration by manual and automatic methods. The cost of checking calibration by the manual method was derived from calibration times only and assumed that procurement and training costs were zero because the manual equipment had already been acquired and training had already occurred. The AMCS cost included its procurement, maintenance, and operator-training costs, as well as its cost derived from time in calibrating meters. Although the computer program for the cost calculation did not proceed in the order listed below, the result was the same as if the following steps had been performed:

1. The average time to calibrate each of the 15 meter types by the manual method was obtained from the first part of the experiment.
2. Using the number of meters of each type at a given site multiplied by the number of calibrations required, the total number of hours required at that site in one year by the manual method was calculated by using these average times.
3. A labor cost per productive hour at that site was calculated by dividing the annual labor cost (that is, wages, fringe benefits, and related overhead cost) of one operator by the number of productive hours of work expected for a typical operator. The number of productive hours was taken as two-thirds of the total hours worked.
4. The annual cost of manual calibration at a site was calculated by multiplying the total number of hours (Step 2) by the cost per productive hour (Step 3).
5. The average time to calibrate each of the 15 meter types by one of the competing AMCSs was obtained from the second part of the test.
6. The cost of calibrating the 15 meter types at a given site by that AMCS was calculated by the same procedure as in Steps 2, 3, and 4, except that the labor cost specified was one of a lower pay grade than that specified for manual calibration. The reason is that manual calibration typically requires operators with higher skills than does calibration using automated equipment.
7. For that AMCS at a given site, the annual cost saving was calculated by subtracting the cost of calibration with the AMCS from the cost of calibration with manual equipment.
8. The present value of these savings was calculated by discounting the stream of savings for 4.5 years at a discount rate of 10 percent.
9. For that AMCS, the savings at each site considered were calculated by repeating the above steps for all 27 sites. Then the savings at each of the seven sites with the largest savings

were added. The reason for this step is that seven AMCSs were to be procured, and presumably they would be assigned to the sites where they would produce the greatest savings.

10. The investment, maintenance, and training costs of that AMCS, as quoted by the vendor in his response to RFP No. 2, were subtracted from the present value of the savings, to give a net cost advantage. The net cost advantage, in principle, could be positive or negative.
11. Steps 5 through 10 were repeated for each of the other two competing AMCs.
12. The AMCS with the greatest net cost advantage received the highest score, and the other two AMCSs were assigned correspondingly lower scores.

The actual computer program was more complicated because it also computed the savings from procuring alternative numbers of AMCSs.

Conclusions About the Cost Model

The calculation of annual hours saved by each AMCS produced a sound number. However, certain improvements could have been made in the procedure for converting these savings to dollar amounts. Specifically, (a) the model assumed that the useful life of an AMCS was 4.5 years, a period too short; (b) no allowance was made for an increase in wage rates during this period; (c) no allowance was made for a possible learning curve effect; and (d) no allowance was made for the reduction in inventory that would be possible with a faster turnaround time for calibration. (The Army told the committee that assumptions about the useful life, wage rate, and zero learning curve were required by Army regulations. The Army's stated objective, however, was to procure the most cost-effective equipment. The committee did not evaluate the degree to which regulations may have diminished the quality of information used to attain this objective.) Thus, the savings over the life of the AMCS was likely understated. Also, the training cost used in the model was only the training cost charged by the vendor; it did not include the labor cost of the trainees. These factors may or may not be significant. They should be taken into account in using the results of the cost model.

The equipment and maintenance costs incorporated in the model were valid. The value assumed for the ratio of productive hours to total hours was less certain because of lack of field data.

Even disregarding the inadequacies mentioned above, the net cost advantage calculated in the model did not measure the actual net cost advantage. The savings calculated were only those associated with calibration times; they did not include any resulting from others of the 18 characteristics listed in Chapter 2. This narrow definition of savings was intentional, because the cost model was restricted to savings information derived from the experiment. It was warranted, provided the effect of other factors was taken into account in other parts of the evaluation.

In addition the cost model assumed that the AMCSs would be available for calibration use for enough time each year to discharge the established calibration workload. Accordingly, workload is the limiting factor on cost savings. The assumption appears reasonable for the actual calibration workloads of about 2,000 calibrations per site per year and actual calibration times of about 0.3 hours per calibration. The cost model does not address the case where the availability of the AMCS is the limiting factor on cost savings. Such a limitation would appear if either the workload or the AMCS downtime were to increase greatly over current estimates of these quantities.

Calibration Stability of the Automated Meter Calibration Systems

The next element in the vendor evaluation model rates the ability of an AMCS to retain its own calibration over time. In particular, if no recalibration was needed during the 45 days of the test, a score of 1.00 was awarded; the need for one recalibration produced a score of 0.5; and for two or more, a score of 0. The scoring was based on the expectation, using a probabilistic model that assumed a constant calibration failure rate, that very few recalibrations would be needed during the test. This expectation results from the fact that an interval of 45 days is small compared with the desired mean interval of 569 days between needed calibrations. For such an interval, a constant hazard model yields the desired probability of 81 percent, used in normal calibration practice, that the AMCS will not need recalibration within the Army's minimum recalibration cycle of 120 days.

The corollary, however, is that the estimate of mean-time-between-loss-of-calibration for a particular AMCS, and hence this aspect of its quality, has poor statistical significance when inferred from a small number of observations. Accordingly, the scoring element has limited validity for discriminating among AMCSs.

Reliability

The third element of the vendor evaluation model pertains to the reliability of the AMCS. Zero, one, two, and three or more failures during the 330 hours of testing received respective scores of 1.00, 0.75, 0.34, and 0. Again, the scoring was based on the fact that few failures would occur because the mean-time-between-failures (MTBF) of an AMCS would probably be larger than 330 hours. For example, the desired MTBF was 536 hours. For reasons similar to those given for the preceding variable, this scoring element also provides information of limited validity.

Additional Meter Calibration Capability

The fourth element in the vendor evaluation model rates the additional capability of an AMCS. An implicit assumption is that this additional capability has statistical properties similar to those associated with the 15 test meters. This element was scored by counting, rather than observing, the tasks that an AMCS can perform, compared with the number of tasks required to calibrate the 57 untested meters listed in RFP No. 1, Table 4 of Attachment A. This factor should not be mistaken for the demonstration, allowed under Section H-5 of RFP No. 1, of additional capabilities to support any test, measurement, and diagnostic equipment (TMDE) within the Army. The weight of 8 percent assigned to this element seems small relative to its potential economic and functional value in the field.

Diagnostic Ability

The fifth element in the vendor evaluation model places a value on the diagnostics of an AMCS. As stated in RFP No. 1, diagnostics is the capability to perform tests, adjustments, and diagnoses of faulty TMDE. This characteristic as tested was the ability of the AMCS to diagnose faults placed in only seven sample meters of the same make and model, although the results were applied to the universe of the Army's TMDE. True diagnostics requires the use of an indicating instrument--typically an oscilloscope or a digital voltmeter. Under RFP No. 1, such an instrument was not a required component of the AMCS, and if it had not been provided by the vendor, the Army would have had to provide it for field use. The scoring system does not appear to compensate in any way for the presence or absence of an indicating instrument for diagnostic purposes in the AMCSs.

Operator Opinion Survey

The limited importance given to the operator opinion survey by the Army reflects the general inadequacies of such surveys. On occasion, self-report evaluative data may have to be used because no objective data can be obtained; and the risk of having no data is often greater than the risk of having spurious data. In the present case, objective data on combined man-machine performance are available in the time measures for calibrating meters with the three AMCSs. These data should have been used in place of questionnaire ratings and rankings of the AMCSs as a better measure of the satisfactoriness of the automated equipment as used by human operators.

Although the survey was designed by specialists, various circumstances make it probable that the survey's results were distorted by misunderstanding or bias. The questionnaire had no pretest, a standard and essential safeguard, other than an exchange of critiques between its two authors and testing with the manual system before use with the AMCSs. The idiosyncratic judgments of one or a few individuals out of the small number of respondents may have distorted the collective view. Extraneous factors could have entered the operator's judgments to create the so-called halo effect. For example, an operator's ratings of performance may have been influenced by the amount and type of training he received, by a system's capabilities for functions other than calibrating the 15 meters, or by friendly associations with representatives of vendors. The construction of questionnaire items could create misunderstandings and misinterpretations for the reasons that follow.

All but one of the rating items incorporated 9-point scales, rather than 5-point or 7-point scales, making discrimination difficult. (See Appendix F.) The rating scales were bipolar, though it is questionable whether the subjective distance from the center to the end is equivalent for both sides of the scale. This issue is particularly raised by the "convenient-inconvenient" and "satisfactory-unsatisfactory" items of the test. Criteria should be performance-centered, as in "easy-difficult," rather than self-centered, as in "satisfactory-unsatisfactory"; extraneous factors are likely to bias expressed feelings about some equipment even more than to bias performance judgments. In any case, "satisfactory" is ambiguous. Satisfactory for what? "Convenient" is also ambiguous. Item 6 referred to "knobs and switches" as if covering all control devices, but other such devices (for example, keyboards) were used with cathode ray tube displays.

The diagnostic data in items 5, 6, 7, and 14 might provide some useful information for designers, but not for the evaluation. The training data in items 9 through 12 could help explain any differences between systems in the objective results. For example, insufficient training might account for one system doing less well than another. If the training provided by the manufacturers is regarded as part of the procurement, a favorable rating for training might contribute a plus to the evaluation. Contrariwise, an unfavorable rating for training might reveal performance potential that had not been realized in the test.

USEFULNESS OF THE METHODOLOGY

The previous discussion shows that the cost model was the most elaborately analyzed factor entering the total evaluation. Even so, the cost model was shown to be susceptible to certain improvements.

Five other factors were included in the evaluation, but the information about them obtained from the test was shown to yield information of limited validity about their true characteristics. Furthermore, the weights assigned to each were judgmentally determined. The attempt to incorporate these factors into the model was commendable, given the constraints of the project. Nevertheless, the decisionmaker should understand the limitations of the attempt. An example of the inadequacy of the information is the rather sketchy demonstration of diagnostic capability, using only seven samples of one type of meter.

In addition, the evaluation model did not incorporate a number of the 18 characteristics listed in Chapter 2 that also relate to performance and life-cycle savings of the AMCSs. Examples are utility in adjusting meters, useful life, amount of downtime, and potential additional capability to support other Army TMDE (demonstrated at the option of the vendor but not scored). Those factors whose importance can be estimated might well be given an appropriate weight and incorporated into a formal evaluation model. The decisionmaker should also give some consideration to the effect of relevant, but unevaluated, factors. There are several ways to do this:

1. If the effect on cost-effectiveness is judged to be approximately the same for all vendors, the factor can safely be disregarded.
2. If a factor increases the cost-effectiveness of an AMCS that has the highest score in the evaluation model, that factor can be disregarded. Inclusion of the factor merely increases the advantage of a vendor and does not change his ranking. If, for example, the AMCS with the highest rating performed the calibration function in the shortest time, and if it is judged that this AMCS could also perform the adjustment and management information functions in the shortest time, then including these times as a factor would increase the cost advantage of this particular AMCS, so its rating would be even higher than that reported.
3. If, however, the influence of the unevaluated factors might affect the rankings, the decisionmaker is faced with a problem. For example, the AMCS with the second highest rating may have the greatest operational savings (as measured by calibration time) but may have investment costs so much higher than that of the highest rated AMCS as to offset the savings of the latter system. It is quite possible, for example, that the investment cost for a second rated AMCS is high because it has capabilities that a lower cost AMCS does not possess. In this circumstance, the decisionmaker must either use judgment about the importance of these capabilities, or request more information before reaching a decision.

In summary, the methodology here described is probably useful in displaying gross differences among AMCSs based on the rather narrow function of meter calibration alone. Small differences would be hard to distinguish with confidence. It is the prerogative of the Army to emphasize the measure of meter calibration. The committee's view, however, is that the Army could realize greater benefits in the long run by more fully taking into account other characteristics of AMCSs.

A more reliable ranking, to satisfy the stated requirements of the Army, would probably have resulted from broadening the purely formalistic methodology to include other factors of importance.

5

ASSESSMENT OF THE ARMY'S REPORT OF THE TESTS

The Army's final report (Appendix C, Item 9) on its evaluation of the three automatic meter calibration systems (AMCS), together with a report limited to a description of the test itself (Appendix C, Item 8), is a summary of the Army's test design, data, and analysis. Comments assessing the final report are arranged here, for the convenience of the reader, under the headings that correspond to the main section titles in the Army report.

EVALUATION METHODOLOGY

The Army's report describes the experimental procedure and outlines the main features of the evaluation methodology. The committee has reservations, however, about the conclusiveness of the evaluation in view of several uncertainties inherent in the test results and their evaluation.

The weights applied to the variables of the scoring model (the cost savings of an AMCS, the number of recalibrations it required, its dependability, its ability to calibrate additional meters, its diagnostic capability, and the operator opinion survey results) were arbitrary in the sense that no attempt was made to apply a common unit of value or utility before combining them. Presumably, each variable could be associated with an economic value or other consistent expression of utility by the exercise of appropriate judgment. In an effort to achieve a purely "objective" score, such judgment was not attempted. As a consequence, the results may undervalue or overvalue one or another of these variables, and therefore no monetary value should be imputed to elements of the scoring model other than the cost element. Furthermore, as the Army concedes in its report, a number of intangible or otherwise unmeasurable benefits of automation are of necessity neglected in the evaluation.

Another source of uncertainty in the results is the choice of a factor relating the labor productivity in test experience to that expected in the field. The outcome of the evaluation is rather sensitive to the choice of a productivity factor; in an analysis of this sensitivity, the Army used a range of factors from 0.4 to 0.8 about its best estimate of 0.66.

A third source of uncertainty is introduced by specifying a 54-month amortization period for AMCSs, when a longer period would be more typical of actual experience in industry and the military. The result of specifying such a short period, of course, is to overemphasize capital cost in the evaluation model. The assumed lifetime of an AMCS of 54 months is probably too short, given that the automated calibration systems in use by the U.S. Air Force and U.S. Navy have a useful economic life of some 10 years.

The uncertainties introduced by the arbitrary weighting procedure, the productivity factor, and the 54-month amortization period all indicate the need for caution in applying the evaluation's results. These results can guide selection of a calibration system, but used without additional considerations they are incomplete and possibly misleading.

DATA SUMMARY

The "Data Summary" section of the Army report consists of tables summarizing the data used in the cost comparison evaluation, the equipment performance evaluation, and the operator opinion survey.

Tables of the times to calibrate each of the 15 different meters are presented, arranged by operator and by trial, for the manual system and for each AMCS. The AMCS times presented in these tables included manual augmentation times for calibration functions, or parts of functions, which the AMCSs could not perform. Also included in the AMCS times, although not clearly indicated in the report, are the adjustments made to the measured times for (a) measurements not made, which were specified in the technical bulletins prescribing the calibration points, and of which the AMCS was judged capable; and (b) superfluous measurements made. In the former case, time was added to the measured calibration times, to account for the additional steps; in the latter, time was deducted.

This section also presents the average time that each vendor's system took to calibrate the 15 meters without manual augmentation. It also presents the average manual augmentation components for the AMCS calibration times by meter and by vendor. The fully augmented times for each vendor, totaled or averaged over all meters, might well have been given as a convenience to the reader. No data describing the extent or the magnitude of the adjustments of the AMCS times are given.

The committee believes that a more complete discussion of the modifications made to the time date, particularly the adjustments for superfluous or insufficient measurements, should have been included in the report. Unpublished data made available to the committee by the Army indicated that about 75 percent of all the AMCS time data at the calibration function level (dc current, ac voltage, and so forth) were modified either by taking into account manual augmentation or by adjustment.

The adjustment technique employed would have distorted the results if any portion of the time to set up the AMCS had mistakenly been included in the adjustment. For a portion of the modified data, the committee was concerned that this situation might have occurred. However, the Army researched the original data, isolated 85 percent of the data that might have been subject to mistaken treatment; and reported that whether or not these data were included in the calculations made no significant difference.* This analysis quantified the uncertainty that the adjustments introduced, permitting the committee to conclude that such data had no serious consequences for the results of the evaluation.

Because the time data were extensively modified, it would have been advisable to have fully documented the criteria governing the augmentations and adjustments. An examination of such criteria revealed no means of ensuring the integrity of the adjustment or of giving full credit for measurement tasks similar, but not identical, to required tasks. However, discussions with the Army personnel involved in the data manipulation task led to reasonably high confidence on the committee's part that the data manipulation task was performed properly.

The Army report presents a table of the number of tasks which each vendor could perform for each of 72 meters (representing a worldwide population) compared to the number of tasks required, using a definition of tasks that was at variance with the definition used in the experiment. This information was used in the evaluation to indicate the AMCSs' ability to calibrate Army meters not used in the experiment. To bring the two diverse definitions into ostensible conformity, a correction factor peculiar to each vendor was derived, as is explained in the section of the Army report, "Evaluation Methodology." The committee does not believe that the use of the correction factor was justified, because it merely mixes the disparate definitions; but the committee is satisfied that its use has only a small effect upon the results of the evaluation.

The performance results relating to calibration stability, dependability, and diagnostic capability are listed for each AMCS; and the related data are presented in a straightforward manner.

The results of the operator opinion survey are presented by question and by operator, and the scores are summarized. The presentation of these data is complete and detailed.

---

*The Army has undertaken to document these findings as an appendix to its report of the tests.

The Army's report showed that the design of the experiment to observe the times required to calibrate meters satisfied the conditions necessary for a formal analysis of variance. Such an experimental design was necessary to ensure that variables in the experimental situation were recognized and controlled. Familiarity with the statistical variation of calibration times among meters, AMCSs, and operators would give an indication of the experiment's important variables and their interactions, the sensitivity of discrimination of random and fixed effects, and the statistical power of the experiment. Although the Army's report does not contain the results of such an analysis of variance, one was performed independently by the Army Materiel Systems Analysis Activity (Appendix C, Item 9), and the results were summarized orally for the committee. This analysis forms a valuable adjunct to the Army's report.

Three manual systems were used in the experiment to provide baseline calibration times for use in the manual augmentation of the data. It would have been informative to have designed the experiment to allow detection of variance among these systems, to test whether they were from the same population. However, the experimental design did not allow the variance of these systems to be identified. The committee does not suggest that the experiment was impaired by this fact; only that, in hindsight, a chance to gather useful information was missed.

EVALUATION RESULTS

The Army's report of the tests discusses the results of its evaluation under several headings. The first is "Cost Advantage." The methodology estimating the net cost advantage resulting from the combination of operational savings offset by capital investment is straightforward.

Results of the analysis depend importantly on the values assumed for workload, for the useful economic life of the AMCSs, and for the ratio of productive hours to total paid hours of calibration personnel. These values cannot be fixed with certainty. As a result, the cost advantages cited should be understood not to be precise, but rather to fall within a more or less uncertain range about the results given by the formalism. The extent of this uncertainty is more fully explored in an analysis of the sensitivity of the formalism to variations in the input data and weights.

The Army's report then combines the individual scores for cost advantage, the several attributes of AMCS performance, and the survey of the opinions of the operators according to weights previously assigned. This formalism is useful in displaying gross differences among AMCSs, as an aid to a decisionmaker. However, the overall scores should not be interpreted as definitive figures of merit, because they are not really the sums of commensurable quantities, as was pointed out above.

A highly useful part of the Army's report is its sensitivity analysis. The relative ranking of Vendor 2 and Vendor 3 is shown to become about equal if the assumed economic lifetime is extended from 54 months to 90 months and the assumed ratio of productive to labor hours is lowered from 0.66 to 0.40. The sensitivity analysis shows that plausible variations in other scoring factors taken one at a time would not be enough to drive the overall scores of Vendor 2 and Vendor 3 to equality or reversal. However, simultaneous, relative changes in the same direction of the order of 0.1 (out of a maximum score of 1) in several of the scores would be enough to eliminate the existing difference of 0.054 between Vendor 2 and Vendor 3 in the overall weighted scores (which lie in the neighborhood of 0.8). Changes of such magnitude are plausible, as shown by the sensitivity analysis itself.

From the beginning, the test was designed to evaluate various AMCSs for augmenting existing manual instrumentation rather than to evaluate a choice between procuring AMCSs or not. A result of the cost-model portion of the evaluation, however, was the conclusion that any AMCS procurement alternatives allowed by Request for Proposal No. 2 would result in negative cost savings, that is, losses. The alternative of not buying any AMCSs at all was the most attractive choice on the basis of considering cost alone; at least no money would be lost. In view of this result, the Army's report undertook to score the "buy nothing" alternative (Appendix C, Item 9, page 54), giving it a perfect score of 1.000 for savings and zero scores for performance and operator opinion. It appears that these zero scores were wrongly assigned because if nothing is bought, the manual system remains. The manual system does exhibit, to some degree at least, the performance characteristics specified by the evaluation model, and does enjoy some degree of favorable operator opinion. Accordingly, the evaluation scores of the "buy nothing" alternative are surely understated; but there is no rationale for properly scoring it. Therefore, the numerical formalism should not be relied on to decide to procure AMCSs or not to procure them. Judgment is also required.

The Army's report concludes with a qualitative discussion of additional considerations that could not be given quantitative treatment within the constraints of time and funding. This section is perhaps the most useful one to a decisionmaker. The report states:

> ...[T]here was a limit to the indication of the potential benefits of automation that could be defined by this program. Many potential benefits that would accrue from introducing AMCS[s] were not examined during the analysis....
> ...[O]nly a small part of the cost savings potential associated with the use of automated meter calibrators has been explored through this program....
> The areas just discussed by no means exhaust the possibilities. Indeed, they probably only scratched the surface of potential benefits. Nevertheless, they serve to illustrate that the Fixed Site AMCS evaluation program provides only a conservative indication of the cost-effectiveness associated with introducing automated calibrators to the Army.
>
> --Appendix C, Item 9, pp. 58, 60.

ASSESSMENT BY THE COMMITTEE

The committee's review of the Army's report indicates that the Army's evaluation project produced much valuable information. However, its results cannot be relied on to indicate the most cost-effective choice of calibration systems that score closely together.

It seems apparent that the project did distinguish the system of Vendor 1 clearly from those of Vendor 2 and Vendor 3 in terms of cost-effectiveness. The score of Vendor 1 is sufficiently lower than those of the other two vendors to make it unlikely that the difference could be due to the uncertainties inherent in the data and their manipulation.

However, the relative cost-effectiveness of the systems offered by Vendor 2 and Vendor 3 cannot be so clearly resolved by the closely spaced scores the model yielded for these vendors. This result, in the committee's view, stems from the uncertainties, and lack of statistical estimates of their magnitude, in the observed data; the model's sensitivity to the assumed productivity factor and economic lifetime of the AMCS; and the benefits not included in the model.

Army selection of one of these two vendors over the other would therefore require additional information, going beyond the evaluation results to a consideration of a number of unquantified benefits and costs, as would a decision to procure any of the AMCSs at all.

6

FINDINGS AND CONCLUSIONS

The principal results of the foregoing assessment can be summarized by five conclusions concerning the design and conduct of the test plan and the evaluation methodology. These in turn are supported by findings concerning their validity, given the purposes they were intended to serve. The conclusions are first stated and then followed by supporting discussion. The discussion is based on the material in the first five chapters and briefly indicates the line of reasoning that leads to the conclusion. A single summarizing conclusion then captures the essence of the committee's views.

The design of the experiment to observe the times for the procedural steps necessary to calibrate meters by the manual and the automated meter calibration systems was sound as judged by established scientific principles, and the experiment was conducted according to carefully prescribed procedures; therefore, the experimental results can be relied on to yield dependable information, subject to uncertainties associated with modifications to the data introduced to achieve commensurability of the test observations.

The times taken to calibrate meters were important inputs to the cost advantage term in the vendor evaluation model. Whereas other inputs to the cost term had fixed values, it was essential to determine the calibration times experimentally to a reasonable level of confidence. Under constraints on time and funding, the experiment chosen was a modified factorial design, with limited replication. Such a design is quite capable of reliable results when the size of the experiment is as large as it was here. The experimental design was adequate for deriving differences among automated meter calibration systems (AMCSs), as well as differences among the other main effects the experiment measured. The design also permitted derivation of the interactions among three of the important main variables.

A number of variables inherent to the experimental situation were adequately controlled through counterbalancing and randomization. The possibility of confounding from extraneous variables was not deemed a serious problem for comparisons among the AMCSs alone. The differences between the experimental situation and the field situation were great enough to warrant some caution as to the degree that the first is assumed to represent adequately the second.

A large portion of the observed calibration time data was modified to produce commensurable data on the AMCSs. The modifications themselves were susceptible to some uncertainties, with the effect of increasing the uncertainty of the time data. However, examination of the probable extent of this uncertainty allowed the committee to conclude that it had no serious consequences for the overall results. (See Chapters 3, 4, and 5.)

Apart from the calibration times, information that was gathered during the test for use in the evaluation methodology--such as that on the reliabilities and calibration stabilities of the automated meter calibration systems, that on their abilities to perform tasks beyond the calibrations measured in the experiment, and that from the survey of operator opinion--had limited validity; although corresponding factors were assigned relatively low weights in the evaluation methodology, the uncertainties associated with them could impair discrimination between two systems with closely spaced scores.

The following five characteristics of an AMCS entered into the vendor evaluation model: (1) its ability to maintain itself in calibration; (2) the number of its hardware failures in a given time; (3) how many tasks it can perform on a specified number of untested meters; (4) its diagnostic capability; and (5) the way its desirable or undesirable features are perceived, as determined by operator survey. Time and funding constraints limited the effort to determine these factors. In all cases, the information was only a very small number of observations of a large universe. Inferences drawn from small samples, of course, have a high risk of not representing well the true values of the population. Individual weights assigned to these capabilities in the evaluation model were low, ranging from 4 percent to 14 percent. Nevertheless, the resulting uncertainties in combination could blur the overall evalution scores. (See Chapters 3, 4, and 5.)

The evaluation model did not include all information pertinent to the merit of automated meter calibration systems because such information could not have been properly quantified for a rigorous, formal model given the constraints of time and funding imposed on the project; therefore, consideration should be given to developing a score for other, qualitative factors in order to optimize the procurement decision.

At least 18 aspects of an AMCS, as listed in Chapter 2, might have been evaluated. Of these, the only statistically significant data that could be collected in the present study were the times that the AMCSs take to calibrate meters. Although some additional information on several other aspects of the performance of the AMCSs was obtained, many of their potential advantages were not studied at all because of the constraints imposed on the project.

Some of the useful features an AMCS may possess, and that might have been accorded some value, include its compatibility with present and future test, measurement, and diagnostic equipment incorporating the Institute of Electrical and Electronics Engineers (IEEE) 488 Instrumentation Bus; its ability to facilitate instrument adjustment and diagnosis; its adaptability to other meters and other test equipment; its capture of data useful in the management of the calibration program; its economic life; and its provisions for assuring a low downtime. (See Chapter 2.)

The economic value of such features was not evaluated, but it could vary widely from one AMCS to another. Even when their benefits are considered in relation to their costs, the extent of the unknown variations in economic and operational utility could perturb the rankings obtained from the formal evaluation model. Omission or inadequate weighting of some of the performance advantages would not change the ranking if the vendor with the highest score was also best with respect to these performance factors. If, however, the vendor ranked second or third offered a system with such superior performance features, proper consideration of them might move one of these vendors into first place. Thus, the decisionmaker should take these factors into account and should obtain additional information, perhaps by the use of panels of impartial experts making independent judgments according to explicit criteria.*

The report of the Army on the cost-effectiveness analysis of fixed-site automated meter calibration systems was generally acceptable, appropriately describing the test plan, the evaluation method, and the evaluation results; the exposition was impaired, however, by lack of a description of the modifications to the calibration time data and the consequences of such modifications.

The report's main shortcoming in exposition was its inadequate discussion of the modification of the raw data. This weakness obscures uncertainties in the evaluation results that had to be separately examined for magnitude and effect. The examination quantified the uncertainty that the adjustments introduced, permitting the committee to conclude that such data had no serious consequences

---

*Statement by Robert N. Anthony: I do not agree with the committee's suggestion for additional scoring, panels, expert opinion, or other information. I believe that the decisionmaker himself should decide whether the Army's report contains enough information for his purpose.

for the results of the evaluation. In addition, if the report had included statistical measures of the variance of the calibration times and possibly of the cost savings, it would have given a useful indication of the uncertainties associated with these quantities. (See Chapter 5.)

Although the report of the Army states that its evaluation model indicates that a particular one of the competing automatic meter calibration systems should be purchased, the caution expressed in that report against reliance solely on formal results is well founded.

The evaluation results, as reported by the Army, score Vendor 3 highest, with Vendor 2 a close second and Vendor 1 a distant third. Given the uncertainties associated with the data, the methodology, and the assumed values of certain inputs, the evaluation is insufficiently precise to distinguish the relative merit of Vendor 2 and Vendor 3. The choice between them should preferably rely on results from additional data on the most important of those characteristics not yet evaluated. In case constraints on the project do not allow the acquisition of additional data, a systematic procedure should be employed to obtain judgmental scores from panels of experts according to specified criteria. (See Chapter 5.)

In the light of the background and stated purposes of the Army's current procurement project, an overall assessment of the test can be expressed by consolidating the foregoing conclusions. The following summarizing conclusion addresses the significance of the test for both the present procurement and the more general future uses mentioned by the Under Secretary of the Army in his Senate testimony. (See Chapter 1.)

The project taken as a whole developed considerable information pertinent to a decision in the present procurement. In particular, the results of the formal evaluation model fairly display gross differences among automated meter calibration systems as an aid to a decisionmaker. If these results are supplemented by systematic scoring of potential, but unmeasured, capabilities of each automated meter calibration system, the Army can very likely make the procurement decision that best serves its requirements for meter calibration and the associated functions that occur in actual operations. However, this particular project suffered from too many constraints to serve as a good prototype for more general procurement evaluations of automated calibration systems.

APPENDIX A

STATEMENT OF TASK*

The overall objective of the Army Calibration Study is to evaluate the U.S. Army's experimental analysis of commercial equipment for calibrating and repairing Army test, measurement, and diagnostic equipment. The study will be initiated with a background briefing at the Army's Calibration Center at Redstone Arsenal, Alabama. Other agencies with specific interests in metrology will also be asked to brief the committee.

The committee will evaluate the following:

1. The experimental design and test procedures to be used by the Army in testing commercial calibration equipment.

2. The methodologies to be used in the analysis of the data.

3. The Army's final report of its tests.

In the conduct of this study the committee will seek the advice and cooperation of other parts of the National Research Council and relevant organizations in government and the private sector. Based on its evaluation, the committee will brief the Army at the various stages of the testing process. The committee will write a report giving its evaluation of the test design and procedures, methodology used in the analysis, and the Army's final report on its test procedures for automatic calibration.

The committee will consist of approximately eight members with expertise in the fields of test design, cost-benefit analysis, calibration, statistical analysis, quality control, human factors analysis, engineering design of calibration equipment, and other relevant disciplines.

---

*Excerpted largely from the National Academy of Sciences proposal, as incorporated into the study contract.

APPENDIX B

LIST OF COMMITTEE MEETINGS

Briefing by representatives of the Army on the background and implementation of the project	April 5-6, 1982 Redstone Arsenal, Alabama
Discussion of prior briefing, progress report by Army representatives, and tentative assessments	April 22-23, 1982 Los Angeles, California
Progress report by Army representatives, discussion of evaluation methodology, and consideration of the report of the committee	May 18-19, 1982 Washington, D.C.
Progress report by Army representatives, discussion to ensure factual accuracy and correctness, and consideration of the report of the committee	June 15-16, 1982 Washington, D.C.
Oral briefing to the Army by the committee on the results of its assessment of the test plan and evaluation methodology	July 19, 1982 Redstone Arsenal, Alabama
Discussion and assessment of the Army's report on the tests	August 23-24, 1982 Washington, D.C.
Oral briefing to the Army by the committee on the results of its assessment of the Army's report of the tests	August 27, 1982 Redstone Arsenal, Alabama

APPENDIX C

LIST OF PRINCIPAL DOCUMENTS REVIEWED BY THE COMMITTEE

1. Letter from John L. Naler, Department of the Army, Office of the Secretary of the Army, to William V. Roth, Jr., United States Senate, with enclosure (containing background material). April 8, 1982.

2. Project Management Plan, Evaluation/Selection of Automated Meter Calibrators for Calibration/Repair of Army TMDE (Test, Measurement, and Diagnostic Equipment, Amendment I. Prepared by Larry H. Johnson, Project Director. February 5, 1982.

3. Solicitation, Offer and Award. Solicitation No. DAAH01-82-R-A193, Issued by the Commander, U.S. Army Missile Command, Redstone Arsenal, Alabama. January 14, 1982.

4. Solicitation, Offer and Award. Solicitation No. DAAH01-82-R-A274, Issued by the Commander, U.S. Army Missile Command, Redstone Arsenal, Alabama. April 5, 1982.

5. Test Plan for Fixed Site Automated Meter Calibration System (Fixed AMCS). U.S. Army Missile Command, U.S. Army Missile Laboratory, Test and Evaluation Directorate, Firing Test Division, Redstone Arsenal, Alabama. March 5, 1982.

6. Methodology for Evaluation of AMCS Proposals. Anon. Undated.

7. Department of the Army Test, Measurement and Diagnostic Equipment Action Team (DATAT). The Deputy Chief of Staff for Logistics. Final Report (the so-called "Bonner Report"). Anon. Undated.

8. Test Report for Fixed Site Automated Meter Calibration System (Fixed AMCS). U. S. Army Missile Laboratory, U.S. Army Missile Command, Redstone Arsenal, Alabama. June 21, 1982.

9. Cost Effectiveness Analysis of Fixed Site Automatic Meter Calibration Systems. U.S. Army Missile Laboratory, U.S. Army Missile Command, Redstone Arsenal, Alabama. August 11, 1982.

10. Crow, Larry H. and Alan W. Benton. Independent Evaluation Report for the Automatic Meter Calibration Study. U.S. Army Material Systems Analysis Activity, Aberdeen Proving Ground, Maryland. To be published.

APPENDIX D

EXCERPTS FROM REQUEST FOR PROPOSAL NO. 1

The full text and exhibits that constitute the Solicitation, Offer and Award, Solicitation No. DAAH01-82-R-A193, informally referred to in this report as Request for Proposal No. 1, are not needed to understand the report. However, selected excerpts are relevant to the definition of the Army's requirements, the specifications of the supplies and services desired, the method of selection of the award, and the method of the subsequent testing of the automated meter calibration systems acquired under the solicitation. Those excerpts are reproduced in this appendix.

DEPARTMENT OF THE ARMY
UNITED STATES ARMY MISSILE COMMAND
REDSTONE ARSENAL, ALABAMA 35898

EXECUTIVE SUMMARY

TO WHOM IT MAY CONCERN:

This Request for Proposal describes goods and services required by the Army to test and evaluate an Automated Meter Calibration System for Fixed Sites (Fixed AMCS). Attachment A describes the hardware, software and self calibration procedures required of the Fixed AMCS. Attachment B contains the requirements for delivery and demonstration of a diagnostic software package. Attachment C describes the contract services for support of government testing.

Proposers are hereby alerted to the fact that a follow-on initial hardware buy is contemplated after an evaluation of the test results and cost effectiveness analysis. See Attachment D. Only contractors awarded a contract under this solicitation will be considered for award for the planned follow-on hardware buy.

Systems proposed which are substantially in excess of system requirements may cause the cost to be above the Government's affordability level and will not be accepted. By submitting a proposal, the contractor agrees that the Government is the sole judge of whether an offered system is either substantially in excess or deficient to system requirements and is, thus, non-responsive.

Any written commitment by the contractor within the scope of this contract shall be binding upon the contractor. For the purpose of this contract, a written commitment by the contractor is limited to the proposal submitted by the contractor, and to specific written modifications to the proposal. Written commitments by the contractor are further defined as including any warranty or representation made by the contractor in a proposal as to hardware or software performance, total system performance, and other physical, design or functioning characteristics of a machine, software package or system. Such warranties or representations made by the contractor with relation to equipment to be tested will be used by the test agency as a baseline for all tests.

A firm fixed price contract is contemplated as a result of the Request for Proposal DAAH01-82-R-A193.

Attention is directed to paragraph L-15 for specific requirements to be included in any proposal.

This Executive Summary is provided solely as an administrative convenience and is not intended in any way to alter the terms and conditions of this Request for Proposal.

Inconsistancy between this Executive Summary and the Request for Proposal, shall be resolved in favor of those elements in accordance with the Order of Precedence (Apr 1973) clause contained in Section L-1.

Billy F. Perkins

BILLY F. PERKINS
CONTRACTING OFFICER

SOLICITATION, OFFER AND AWARD

3. CERTIFIED FOR NATIONAL DEFENSE UNDER BDSA REG. 2 AND/OR DMS REG. 1 — RATING DO-A2		4. PAGE 1	OF 73
1. CONTRACT (Proc. Inst. Ident.) NO.	2. SOLICITATION NO. DAAH01-82-R-A193 ☐ ADVERTISED (IFB) ☒ NEGOTIATED (RFP)	5. DATE ISSUED 1982 JAN 14	6. REQUISITION/PURCHASE REQUEST NO. FQ175-82
7. ISSUED BY — CODE W31P4Q	Commander US Army Missile Command ATTN: DRSMI-IZA(1)/Lovelady Redstone Arsenal, AL 35898	8. ADDRESS OFFER TO (If other than block 7)	

* In advertised procurement "offer" and "offeror" shall be construed to mean "bid" and "bidder".

SOLICITATION

9. Sealed offers in original and 2 copies for furnishing the supplies or services in the Schedule will be received at the place specified in block 8, or if handcarried in the depository located in Bldg. 4488 until 3:00 p.m. local time 1982 JAN 29 (Date) (See Paragraph L-17)

If this is an advertised solicitation, offers will be publicly opened at that time.

CAUTION – LATE OFFERS: See pars. 7 and 8 of Solicitation Instructions and Conditions

All offers are subject to the following:

1. The Solicitation Instructions and Conditions SF33A, Rev 1-78 edition which is attached or incorporated herein by reference (See L-1)
2. The General Provisions which are incor:,orated herein by reference.
3. The Schedule included herein and/or attached hereto.
4. Such other provisions, representations, certifications, and specifications as are attached or incorporated herein by reference. (Attachments are listed in Schedule.)

FOR INFORMATION CALL (Name & telephone no.) (No collect calls) ▶ Joyce Lovelady/205/876-7557

TABLE OF CONTENTS

(X)	SEC. (1)	PART I – THE SCHEDULE	PAGE	(X)	SEC.	PART II – GENERAL PROVISIONS	PAGE
X	A	Contract Form		X	I	General Provisions	11-19
X	B	Supplies/Services & Prices	3			PART III – LIST OF DOCUMENTS, EXHIBITS AND OTHER ATTACHMENTS	
X	C	Description/Specifications	4	X	J	List of Documents, Exhibits and other Attachments	20
X	D	Packaging and Marking	4			PART IV – GENERAL INSTRUCTIONS	
X	E	Inspection and Acceptance	4	X	K	Representations, Certifications and other Statements of Offeror	21-27
X	F	Deliveries or Performance	4	X	L	Instructions, Conditions and Notices to Offerors	28-33
X	G	Contract Administration Data	5	X	M	Evaluation Factors for Award	34
X	H	Special Provisions	6-10	X		Attachments	35-73

OFFER *(must be fully completed by offeror)*

In compliance with the above, the undersigned agrees, if this offer is accepted within ______ calendar days (60 calendar days unless a different period is inserted by the offeror) from the date for receipt of offers specified above, to furnish any or all items upon which prices are offered at the price set opposite each item, delivered at the designated point(s), within the time specified in the schedule.

16. DISCOUNT FOR PROMPT PAYMENT (See par. 9, SF 33A)

% 10 CALENDAR DAYS	% 20 CALENDAR DAYS	% 30 CALENDAR DAYS	CALENDAR DAYS

17. OFFEROR — CODE — FACILITY CODE NAME AND ADDRESS (Street, city, county, State and ZIP code)	18. NAME AND TITLE OF PERSON AUTHORIZED TO SIGN OFFER (Type or print)	
AREA CODE AND TELEPHONE NO. ▶	19. SIGNATURE	20. OFFER DATE

☐ Check if remittance address is different from above – enter such address in Schedule

AWARD *(To be completed by Government)*

21. ACCEPTED AS TO ITEMS NUMBERED	22. AMOUNT	23. ACCOUNTING AND APPROPRIATION DATA	
24. SUBMIT INVOICES (4 copies unless otherwise specified) TO ADDRESS SHOWN IN BLOCK ______		25. NEGOTIATED PURSUANT TO	10 U.S.C. 2304(a)( ) 41 U.S.C. 252(c)( )
26. ADMINISTERED BY (If other than block 7) — CODE		27. PAYMENT WILL BE MADE BY — CODE	
28. NAME OF CONTRACTING OFFICER (Type or print)		29. UNITED STATES OF AMERICA BY *(Signature of contracting officer)*	30. AWARD DATE

[illegible]

33-120

15 APR 1980

Standard Form 33 Part 1 (REV. 3-77)
[illegible] Prescribed by GSA FPR (41 CFR) 1-16.101

STANDARD FORM 36, JULY 1966
GENERAL SERVICES ADMINISTRATION
FED. PROC. REG. (41 CFR) 1-16.101
EXCEPTION TO SF 36 APPROVED
BY NARS, MAR 1977

CONTINUATION SHEET

REF. NO. OF DOC. BEING CONT'D.	PAGE	OF
DAAH01-82-R-A193	3	73

NAME OF OFFEROR OR CONTRACTOR

ITEM NO.	SUPPLIES/SERVICES	QUANTITY	UNIT	UNIT PRICE	AMOUNT
	SECTION B - Supplies, Services and Prices				
0001	One Automated Meter Calibration System including software and self-calibration procedures in accordance with Scope of Work, Attachment A.	1	ea		
0002	Maintenance Software in accordance with Scope of Work, Attachment B.	1	Lot		
0003	All labor and/or services for Equipment Maintenance, and Training of Personnel and Technical Support in accordance with Scope of Work, Attachment C.	1	Lot		
0004	Two complete sets of Manuals in accordance with DD 1423, Attachment F.	1	se		

NOTES:

1. The contractor's proposal is required to respond to each contract line item to permit full evaluation of both TMDE calibration and maintenance capabilities of the proposed automated meter calibration system. However, the inability to fully perform the requirements for the meters in Table I as specified in Table II will not exempt the contracted equipment from acceptance. Also, inability to fully perform requirements of CLIN 0002 will not exempt the contracted equipment from acceptance, provided the contractor proposal describes the capability to produce software for diagnostic or maintenance operations.

B-2 TYPE OF CONTRACT:

A firm fixed price contract is contemplated as a result of this solicitation.

H-5 SYSTEM DEMONSTRATION

At the contractor's request, with the Government's technical representative's approval, the contractor may be permitted to demonstrate any additional capability or unique features of his hardware/software system within fifteen (15) working days of completion of all planned Government tests.

H-5 (CONT'D)

Such demonstration of any additional capability shall be at the contractor's sole expense. Government technical personnel shall be available only as witness of such demonstration. This demonstration will in no way obligate the Government to evaluate such additional capability. By submitting a proposal, contractor agrees that the decision either to evaluate or not evaluate such additional capability is within the sole discretion of the Government and not subject to the Disputes Clause.

L-15 PROPOSAL INFORMATION

Offerors proposals submitted in response to this solicitation shall include, but is not limited to, the following data:

a. Except where modified herein, the system offered shall be the most cost effective (per his judgment) standard commercial ("off-the-shelf") product of the proposer. All parts, components, and assemblies shall be new, unused, and free from defects, and imperfections which might affect the serviceability and appearance of the finished product(s).

b. The contractor shall be limited to proposing only one Automatic Meter Calibration System (AMCS).

c. The contractor must provide information identifying his "advertised" performance characteristics and environmental limitations for his proposed equipment (including reliability and maintainability).

d. Special Instructions:

The contractor must furnish to the Government catalogs which contain market prices plus copies of invoices indicating sales to the general public with prices annotated thereon or some other indicator that could be concluded as proof that prices quoted under this RFP are, in fact, no higher than the prices quoted to the contractor's most favored customer. If these are not available, the contractor is required to submit a cost breakdown covering material, engineering labor hours and rates, manufacturing labor hours and rates, and applicable overhead rates. The contractor may fill out and provide with his proposal either a Contract Pricing Proposal (DD Form 633) or Claim for Exemption From Submission of Certified Cost or Pricing Data (DD Form 633-7). A proposal submitted without such evidence or offering prices higher than those quoted to the contractor's most favored customer may be considered nonresponsive.

e. The contractor shall be provided storage space for AMCS repair parts at the test site. Requirements must be identified in the proposal.

f. The contractor must identify any limited government rights to proprietary software and include purchase price thereof.

SECTION M - METHOD OF SELECTION

M-1 SELECTION CRITERIA

Systems proposed which are substantially in excess of system requirements may cause the cost to be above the Government's affordability level and will not be accepted. By submitting a proposal, the contractor agrees that the Government is the sole judge of whether an offered system is either substantially in excess or deficient to system requirements and is, thus, non-responsive.

M-2 An offeror must be determined responsible according to the standard in DAR Section I, Part 9 to be eligible for award.

M-3 RESPONSIVENESS

The offeror is cautioned to read and comply with all provisions of this solicitation. To be considered for award, an offer must comply in all material respects with the essential requirements of the solicitation, so all offerors may be equally evaluated.

APPENDIX E

EXCERPTS FROM REQUEST FOR PROPOSAL NO. 2

The full text and exhibits that constitute the Solicitation, Offer and Award, Solicitation No. DAAH01-82-R-A274, informally referred to in this report as Request for Proposal No. 2, are not to understand the report. However, selected excerpts are relevant to the definition of the Army's requirements, the specifications of the supplies and the services desired, and evaluation factors for the award. Those excerpts are reproduced in this appendix.

SOLICITATION, OFFER AND AWARD

1. CERTIFIED FOR NATIONAL DEFENSE UNDER BDSA REG. 2 AND/OR DMS REG. 1 — RATING DO-A2

PAGE 1 OF 70

2. CONTRACT (Proc. Inst. Ident.) NO.	3. SOLICITATION NO.	5. DATE ISSUED	6. REQUISITION/PURCHASE REQUEST NO.
	DAAH01-82-R-A274 ☐ ADVERTISED (IFB) ☑ NEGOTIATED (RFP)	1982 APR 05	FQ 042-82

7. ISSUED BY — CODE W31P40

Commander
US Army Missile Command
ATTN: DRSMI-IZA(1)/Beck
Redstone Arsenal, AL 35898

8. ADDRESS OFFER TO (If other than block 7)

In advertised procurement "offer" and "offeror" shall be construed to mean "bid" and "bidder".

SOLICITATION

9. Sealed offers in original and 4 copies for furnishing the supplies or services in the Schedule will be received at the place specified in block 8, or if handcarried, in the depository located in Bldg 4488 until 3:00 PM (Hour) local time 3 June 82 (Date) (See Paragraph L-17)

If this is an advertised solicitation, offers will be publicly opened at that time.

CAUTION – LATE OFFERS: See pars. 7 and 8 of Solicitation Instructions and Conditions.

All offers are subject to the following:

1. The Solicitation Instructions and Conditions SF33A, Rev 1-78 edition which is attached or incorporated herein by reference (See L-1)
2. The General Provisions which are incorporated herein by reference.
3. The Schedule included herein and/or attached hereto.
4. Such other provisions, representations, certifications, and specifications as are attached or incorporated herein by reference. (Attachments are listed in Schedule.)

FOR INFORMATION CALL (Name & telephone no.) (No collect calls) ▶ Brenda Beck 205/876-4650

TABLE OF CONTENTS

(✓)	SEC.	PART I THE SCHEDULE	PAGE	(✓)	SEC.		PAGE
X	A	Contract Form	1	X	I	PART II GENERAL PROVISIONS: General Provisions	13-18
X	B	Supplies, Services & Prices	2-3			PART III LIST OF DOCUMENTS, EXHIBITS AND OTHER ATTACHMENTS	
X	C	Description/Specification	4	X	J	List of Documents, Exhibits and other Attachments	19
X	D	Packaging and Marking	4			PART IV GENERAL INSTRUCTIONS	
X	E	Inspection and Acceptance	4-5	X	K	Representations, Certifications and other Statements of Offeror	20-27
X	F	Deliveries or Performance	4-5	X	L	Instructions and Conditions and Notices to Offerors	28-33
X	G	Contract Administration Data	6	X	M	Evaluation Factors for Award	34-36
X	H	Special Provisions	7-12	X		ATTACHMENTS	37-70

OFFER *(must be fully completed by offeror)*

In compliance with the above, the undersigned agrees, if this offer is accepted within ________ calendar days (60 calendar days unless a different period is inserted by the offeror) from the date for receipt of offers specified above, to furnish any or all items upon which prices are offered at the price set opposite each item, delivered at the designated point(s), within the time specified in the schedule.

16. DISCOUNT FOR PROMPT PAYMENT (See par. 9, SF 33A)

% 10 CALENDAR DAYS	% 20 CALENDAR DAYS	% 30 CALENDAR DAYS	% CALENDAR DAYS

17. OFFEROR — CODE — FACILITY CODE

NAME AND ADDRESS (Street, city, county, State and ZIP code)

AREA CODE AND TELEPHONE NO. ▶

☐ Check if remittance address is different from above – enter such address in Schedule

18. NAME AND TITLE OF PERSON AUTHORIZED TO SIGN OFFER (Type or print)

19. SIGNATURE

20. OFFER DATE

AWARD *(To be completed by Government)*

21. ACCEPTED AS TO ITEMS NUMBERED	22. AMOUNT	23. ACCOUNTING AND APPROPRIATION DATA

24. SUBMIT INVOICES (4 copies unless otherwise specified) TO ADDRESS SHOWN IN BLOCK ______

25. NEGOTIATED PURSUANT TO: X 10 U.S.C. 2304(a)(10) / 41 U.S.C. 252(c)( )

26. ADMINISTERED BY (If other than block 7) — CODE

27. PAYMENT WILL BE MADE BY — CODE

28. NAME OF CONTRACTING OFFICER (Type or print)

29. UNITED STATES OF AMERICA

BY

(Signature of contracting officer)

30. AWARD DATE

Award will be made on this form, or on Standard Form 26, or by other official written notice.

33-133

15 APR 1980

Standard Form 33 Part I (REV. 3-77)

STANDARD FORM 36, JULY 1966
GENERAL SERVICES ADMINISTRATION
FED. PROC. REG. (41 CFR) 1-16.10
EXCEPTION TO SF 36 APPROVED BY NARS, MAR 1977

CONTINUATION SHEET

REF. NO. OF DOC. BEING CONT'D.: DAAH01-82-R-A274

PAGE 2 OF 7

NAME OF OFFEROR OR CONTRACTOR

ITEM NO.	SUPPLIES/SERVICES	QUANTITY	UNIT	UNIT PRICE	AMOUNT
	SECTION B-1				
	Fixed Site Automated Meter Calibration Systems (Fixed AMCS) in accordance with Attachment A dated 28 Jan 82. This shall include one (1) set of Interface Cables and accessories and one (1) set of Application Software per system as specified in paragraphs 3.4.3.2 and 3.5 of Attachment A. Each contractor shall propose for CLINs 0001, 0005, 0011 and 0012 the same configuration of Fixed Site Automated Meter Calibration System which he proposed in response to DAAH01-82-R-A193				
0001	Fixed AMCS	7	EA		
0002	Training in accordance with Attachment B.	1	LOT		
0003	Necessary Application Software to provide calibration capability, per paragraph 3.5, Attachment A, for additional meters as follows:	1	EA		
	MODEL NO — CALIBRATION PROCEDURE AND SPECIFICATIONS ME 161 — TB-9-6625-1866-50 COHU 202B — TB 11-6625-2555-50/1 JF 8125A — TB 9-4935-525-50-1				
0004	Contract Data Requirements List DD 1423.	1	LOT	NSP	
	OPTION I				
0005	Fixed AMCS	1-20	EA		
	OPTION II				
0006	Maintenance for FY83 in accordance with Scope of Work Attachment C (Contractor's plant). (82 Dec 1 - 83 Sep 30)		UNIT/MO		
	OPTION III				
0007	Maintenance for FY84 in accordance with Scope of Work Attachment C (contractor's plant). (83 Oct 1 - 84 Sep 30)		UNIT/MO		
	OPTION IV				
0008	Maintenance of FY85 in accordance with Scope of Work Attachment C (contractor's plant). (84 Oct 1 - 85 Sep 30)		UNIT/MO		

36-108-17

STANDARD FORM 36, JULY 1966 GENERAL SERVICES ADMINISTRATION FED. PROC. REG. (41 CFR) 1-16.101 EXCEPTION TO SF 36 APPROVED BY NARS, MAR 1977	CONTINUATION SHEET	REF. NO. OF DOC. BEING CONT'D. DAAH01-82-R-A274	PAGE 3	OF

NAME OF OFFEROR OR CONTRACTOR

ITEM NO.	SUPPLIES/SERVICES	QUANTITY	UNIT	UNIT PRICE	AMOUNT
	OPTION V				
0009	Maintenance for FY86 in accordance with Scope Work Attachment C (contractor's plant). (85 Oct 1 - 86 Sep 30)		UNIT/MO		
	OPTION VI				
0010	Maintenance for FY87 in accordance with Scope of Work Attachment C (contractor's plant). (86 Oct 1 - 87 May 30)		UNIT/MO		
	OPTION VII				
0011	Fixed AMCS	1-55	EA		
	OPTION VIII				
0012	Fixed AMCS	1-40	EA		

NOTE I: Contractor must propose on CLINS 0001 through 0012 at the stated quantities.

NOTE II: The Government reserves the right to make initial award of any combination of CLINs 0001 through 0005

NOTE III: The contractor's proposal for Option II through VI shall be based upon a maintenance price per system per month for a quantitiy not less than 7 units and up to and including 27 units.

NOTE IV: Quantities shown for each Option for AMCS (CLIN 0005, 0011 and 0012) are maximum quantities that may be procured. The Government reserves the right to exercise any quantity, (in multiple increments), up to the maximum. On 0005, 0011 and 0012, the offeror shall propose unit prices (and specify quantity price breaks) from one unit up to and including the maximum quantities shown. Variation in quantity is ± 0%.

NOTE V: Unit price proposed for CLIN 0005, shall be no higher than that proposed for CLIN 0001.

NOTE VI: Award may be made to more than one offeror (See M-3 and M-7).

B-2 TYPE OF CONTRACT:

Firm Fixed Price contract(s) is(are) contemplated as a result of this solicitation.

L-12 PROPOSAL INFORMATION

Offerors' proposals submitted in response to this solicitation shall include, but is not limited to, the following data:

a. The Contractor shall propose the same configuration/Fixed Site Automatic Meter Calibration System which was delivered as a result of DAAH01-32-R-A193.

b. It is incumbent upon this agency to establish that any contract awarded as a result of this solicitation be based upon fair and reasonable prices. Therefore, the following information must be furnished in support of any vendor's proposed price.

(1) An established published price list for that material and/or service sold in substantial quantities to the general public. In the event a published price list is furnished in support of your proposed price, a DD Form 633-7 must be submitted as proof that the items(s) are sold in substantial quantities to the general public: or

(2) Pricing information similar to that required on DD Form 633, e.g., direct material, material overhead, direct labor overhead, general and administrative expense, and profit of fee; or

(3) Cost or pricing data as required for DD form 633 (See DAR 3-807.3).

c. Information Requirements

(1) Description of applications for which equipment was designed.

(2) Description of applications for which equipment has been sold to other customers.

(3) Duty cycles associated with operation of the Fixed AMCS.

(4) Manuals or instructions for modification of software and/or procedures (Flow charts, listings, etc.) for use in future "make or buy" decisions for other applications, such as calibration of additional meters or meter repair/diagnostics.

(5) Information and back-up described and substantiating flexibility of the calibrator.

d. Product Function Specification

The contractor shall submit to the Government, as part of his proposal, a Product Function Specification to be negotiated and subsequently incorporated into the contract. The Product Function Specification submitted herein may be modified to reflect the actual results of the Government evaluation tests resulting from Solicitation DAAH01-82-R-A193. This Product Function Specification as modified will become the certification baseline for acceptance of the Fixed AMCS. Any contractor not agreeing to the above condition may be rejected as non-responsive.

The Product Function Specification shall include as a minimum:

a. General requirements of safety, soldering, interchangeability, electrical overload, workmanship and printed wiring.

b. Hardware Performance:

(1) Functions and Parameters

(2) Measurement and Stimulus Ranges

(3) Accuracies

(4). Long-term/Short Term Stability

(5) RMS Noise

c. Environmental Performance; i. e., Temperature Coefficients

d. Electrical Power Requirement

e. System Calibration Interval and Stability

f. Processor and Peripheral Capabilities

g. Software Capabilities

SECTION M - EVALUATION FACTORS

M-1 RESPONSIBILITY

An offeror must be determined responsible according to the standard in DAR Section I, Part 9 to be eligible for award.

M-2 RESPONSIVENESS

The offeror is cautioned to read and comply with all provisions of this solicitation. To be considered for award, an offer must comply in all respects with the requirements of the solicitation, so all offerors may be equally evaluated.

M-3 EVALUATION CRITERIA

Three factors will be evaluated to determine the most cost effective system and to form the basis for selection. These factors are: a. Cost Advantage, b. Equipment Performance, and c. Operator Opinion Survey.

a. FACTOR 1. Cost Advantage.

(1) Cost Advantage associated with using the contractor's proposed Fixed AMCS will be determined by comparing the cost of calibration using the existing manual system to the cost of performing the same calibration tasks using an augmented manual/automated system. This comparison will be done using a scenario that considers the established workload for 27 CONUS managed (i.e., CONUS, Alaska, Hawaii) fixed sites. Workload will be projected over a 54 month period and will include only the 15 models of meters specified in Attachment A, Table 1. Cost elements to be included are cost of AMCS acquisition (unit price for baseline quantity), cost of maintenance, cost of training, and cost of operation. The first three cost elements will be obtained from contractor proposals and the last cost element will be determined by using operational test data from Phase I tests (test with equipment resulting from Solicitation DAAH01-82-R-A193). Evaluation will be conducted utilizing constant FY83 dollars with Fiscal Years discounted appropriately and utilizing actual operator manhour rate for each fixed site. A favorable cost advantage exists when total cost of ownership and operation of the augmented manual/ automated system is less than the total cost of operation of the existing manual system. Acquisition, maintenance support and training cost for manual systems will not be included in the evaluation.

(2) The government recognizes that equipment from different offerors may have different performance characteristics and, therefore, may have different levels of cost advantage at each of the fixed sites (depending on workload and distribution of meter types). Therefore, more than one award of 7 each may result.

b. FACTOR 2. Equipment Performance.

Performance characteristics not explicitly included in Factor 1 will be evaluated under this factor. These characteristics include the frequency of recalibration of the Fixed Site AMCS, its dependability, and its potential in terms of its ability to calibrate additional meters as set forth in Table IV of Attachment A as well as its potential diagnostic capability. This evaluation will be based on data collected during the Phase I test and information required in the response to this RFP.

c. FACTOR 3. Operator Opinion Survey

Operator Opinion Survey inputs will be evaluated to include adequacy and ease of training and interaction of personnel with equipment (ease of operation).

SUMMARY

The overall cost effectiveness of each offer will be calculated by weighting and combining the results of the Factor 1, 2, and 3 evaluations. This will be the basis for selection.

COST ADVANTAGE WILL BE MORE SIGNIFICANT THAN ALL OTHER FACTORS COMBINED. FACTOR 2 WILL BE MORE SIGNIFICANT THAN FACTOR 3.

M-4 RESIDUAL VALUE

For the purposes of this CEA evaluation, the residual value of equipment and operating software, will be zero (0).

M-5 OTHER EVALUATION CRITERIA

Discounts. Proposer shall furnish the government a schedule of reasonable discounts applicable to the following:

1. Prompt Payment Discounts; and

2. Any other special discounts the proposer desires to offer.

All Discounts will be considered in the CEA evaluation.

M-6 EVALUATION OF OPTIONS

All proposers are cautioned to be cognizant of the fact that the proposed prices for all maintenance options associated with the maintenance of the basic award quantity will be utilized in the CEA evaluation. Proposed prices for options for increased quantities will not be considered in the CEA evaluation. Evaluation of the options does not obligate the Government to exercise any options.

M-7 AWARD OF CONTRACT

The contract(s) will be awarded to responsible offeror(s) whose offer(s), conforming to the solicitation, will be most advantageous to the Government, price and other factors considered.

M-8 FIXED PRICE OPTIONS

Proposers must offer all options solicited in Section B of this Request for Proposal in order to be responsive. To be considered for award, all options submitted must be proposed on a fixed price basis. See, however, paragraph H-6 entitled "Economic Price Adjustment".

APPENDIX F

METHODOLOGY FOR EVALUATION OF AMCS PROPOSALS

For a full understanding of this report, it is important that the reader have access to a description of the methodology that the Army used to evaluate the vendor proposals for automated meter calibration systems. A description of that methodology was provided to the committee by Army representatives and is reproduced in this appendix.

METHODOLOGY FOR EVALUATION OF AMCS PROPOSALS

Proposed AMCS equipment will be evaluated using three factors. These are (1) Cost Advantage, (2) Equipment Performance, and (3) Operator Opinion Survey. The following is a description of the methodology that will be used to determine a score for each proposer on each of the factors, and to combine the scores on each factor to obtain a preferred approach for satisfying the Army's requirement.

A. Description of Factors:

1. Cost Advantage. The cost advantage to the Army associated with using the contractor's proposed AMCS will be determined by comparing the cost* of calibration using the existing manual system to the cost of performing the same calibration tasks using an augmented manual/automated system. This comparison will be done using a fixed scenario that consists of the established workload for 27 CONUS managed fixed sites. This workload will be projected over a 54 month period and will include only the 15 models of meters specified in the RFP. A cost advantage will exist for the proposed AMCS whenever the acquisition cost, training cost, maintenance cost, and operating cost for the AMCS is less than the operating cost of the manual system. Since the manual system is already in place, its acquisition cost and training cost are considered a washout. Also, since the manual system will be used to perform any tasks which cannot be performed by the AMCS, it will always exist and its maintenance cost is considered a washout. The potential cost savings at each site will be calculated for each vendor's AMCS. Calibration tasks that the AMCS cannot perform will be assumed to be performed manually and the cost to automatically perform the calibration will be increased by the cost to manually perform these additional tasks. Present value in FY 83 dollars will be used for all costs.

The basic quantity of AMCS for each vendor will be assigned to those sites for which the greatest savings potential exists in a manner that will result in the maximum possible savings for the Army. As an example, suppose the basic quantity of AMCS from any vendor is 5 sets, and that calculation of the costs of performing the calibrations manually and with AMCS, and the resulting savings, are as shown in Table I for a hypothetical 10 sites. Note that the last three columns, which indicate the savings potential associated with locating vendor v_i equipment at site j, is obtained by subtracting the cost of performing the workload at that site with AMCS from the cost of doing the same job manually. Also note that negative savings are included as was done for v_1 at site 2 and v_2 at site 3. The circled values in the savings column indicate the assignment of each vendor's AMCS to those sites that will maximize the savings potential

*Cost groundrules are attached.

for the basic quantity of 5 units. The sum of these circled quantities for each vendor gives the maximum savings possible from this basic quantity. It is clear that the AMCS of vendor 1 yields the greatest potential savings.

TABLE 1. Hypothetical Costs and Savings

Site	Cost				Savings		
	Manual	V_1	V_2	V_3	V_1	V_2	V_3
1	100	73	87	93	27	13	7
2	7	17	5	7	-10	2	0
3	44	32	45	40	12	-1	4
4	34	28	30	26	6	4	8
5	46	31	41	44	15	5	2
6	68	57	50	63	11	18	5
7	38	31	27	25	7	11	13
8	65	50	60	59	15	5	6
9	76	68	71	70	8	4	6
10	70	56	64	53	14	6	17
					83	53	51

If the buy quantity was 10 rather than 5, then the savings for all 10 sites would be summed for each vendor, with the result that the total savings potential would be 105 for vendor v_1, 67 for v_2, and 68 for v_3.

In the event that more than a basic quantity of 5 AMCS is procured, the potential savings may be larger if a mix of units is procured from different vendors. For the example case of 3 vendors, there are 3 distinct pairings, as shown in Table II. The assignments can be made by picking the vendor with the greater savings at each site, as indicated by the circled values in the table.

TABLE II. Hypothetical Savings, Multiple Awards

Site	Pairings					
	V_1	V_2	V_1	V_3	V_2	V_3
1	27	13	27	7	13	7
2	-10	2	-10	0	2	0
3	12	0	12	4	-1	4
4	6	4	6	8	4	8
5	15	5	15	2	5	2
6	11	18	11	5	18	5
7	7	11	7	13	11	13
8	15	5	15	6	5	6
9	8	4	8	6	4	6
10	14	6	14	17	6	17
Total Savings	128		126		92	

In this example, the pairing V_1 V2 gives the greatest potential savings, with 7 units procured from vendor v_1 and 3 from v_2.

If a procurement of at least 5 AMCS from each vendor is a constraint, then this can be accommodated by modifying the result obtained in Table II in such a manner that the reallocation results in the smallest reduction in total savings potential. This is accomplished by changing the assignment at site 4, for a savings loss of 2, and at either site 7 or 9 for an additional loss of 4, for the v_1v_2 pairing, at sites 9 and 6 for a total loss of 8 in the v_1v_3 pairing, and at site 8 for a loss of one in the v_2v_3 pairing.

The preceding description covers the possible procurement alternatives and the allocation of AMCS to sites in a manner that maximizes the potential savings to the Army.

2. Equipment Performance. The equipment performance characteristics which will be evaluated under this factor are the frequency of recalibration of the fixed site AMCS, its dependability, and its growth potential in terms of its ability to calibrate additional meters as well as its diagnostic capability. Data on which to base this evaluation will be obtained from the Phase I tests and from the response to the RFP. The score for equipment performance will be a value on the interval [0, 1.0] which will be obtained by multiplying the scores for each of the above subfactors by a weighting factor on the interval [0, 1] and adding. The method for scoring each subfactor is defined below.

a. Frequency of Recalibration. All AMCS will be tested after receipt to verify the precision of measured parameters using calibration standards. These tests will be conducted by personnel of the Army Metrology and Calibration Center, and verified by representatives of the National Bureau of Standards. At the end of each test cycle, each AMCS will be retested against the calibration standards. From these tests, the number of times any vendor's AMCS requires recalibration will be determined. (Recalibration required as a result of a repair action will not be scored in this subfactor; i.e., it will not be included in this count.) The normal standard for calibration equipment is that 81% of the items will not require recalibration before the end of the recalibration interval. The required recalibration interval for the AMCS is 120 days (RFP DAAH01-82-R-A193, Attachment A). Using a constant hazard model (exponential failure rate), these conditions imply that the mean time between calibrations for the AMCS should be at least 569 days. Each AMCS will be used for approximately 45 days during the test program. If the hypothesis that the mean time between recalibrations exceeds 569 days is true, then the probability of more than one recalibration being required during the test program is less than .02. If any recalibrations are required, this fact provides a substantial amount of information about the uncertainty of the mean time between recalibrations. Consequently, the following scores will be assigned:

# Recalibrations	0	1	2 or more
Score	1.00	.50	0

b. Dependability. The number of failures of each AMCS during the training and test periods will be recorded. While there is no stated requirement for MTBF in the RFP, it is desirable that the reliability of the AMCS be at least as good as that of the manual system. The composite MTBF of the calibration equipment in the manual system is 537 hours, using a constant hazard model. Each AMCS will be tested for approximately 330 hours. If the MTBF of the AMCS is at least 537 hours, then the probability of more than two failures occuring during the test period is less than .025. As with recalibration frequency; any failures provide significant information about the uncertainty associated with the MTBF. Hence, the following scores will be assigned:

# Failures	0	1	2	3 or more
Score	1.00	.75	.34	0

c. Ability to Calibrate Additional Meters. The ability of each AMCS to calibrate meters other than the 15 meters in the test program will be evaluated by comparing the technical requirements for calibration of the remaining 57 meters in Table IV of RFP DAAH01-82-R-A193, Attachment A, with the capability of the AMCS to perform these calibrations. The score for each AMCS will be the percentage of the total required number of calibration actions that can be performed by the AMCS. That is, letting NC_i be the number of calibration tasks on meter i, the total number of tasks for the 57 meters would then be

$$NC_T = \sum_{i=1}^{57} NC_i \ .$$

If NC_{iv} is the number of calibration tasks on meter i that can be performed by the AMCS of vendor v, then the score on this subfactor will be

$$S_v = (\sum_{i=1}^{57} NC_{iv}) \ / \ NC_T$$

d. Diagnostic Capability. The score for diagnostic capability will be based on the capabilities demonstrated during the vendor demonstration required by contract line item 2 of RFP DAAH01-82-R-A193. For this demonstration, each vendor will be presented with 8 meters of the same type, each of which will have a specific fault. The nature of the fault will not be known to the vendor. Multiple faults in a single meter may be set up, and some meters will have no faults at all. The vendor was told which meter would be used for this demonstration in RFP DAAH01-82-R-A193, and presumably has developed and implemented software and hardware that will facilitate the diagnosis of its faults. Let NF be the number of faults in the sample of meters, and let ND_v be the number of these faults that are successfully diagnosed by the AMCS of vendor v. ND_v will be decreased by one for every false alarm. Then the score on this subfactor will be

$$S_v = ND_v \ / \ NF \ .$$

The weights used to combine each of the above equipment performance subfactors shall be as follows:

a.	Frequency of Recalibration	.35
b.	Dependability	.35
c.	Ability to Calibrate Additional Meters	.20
d.	Diagnostic Capability	.10

The method for calculating the score for each vendor on this factor is illustrated in Table IV.

Table IV. Equipment Performance Score

	Wt	Score			Wtd Score		
		V_1	V_2	V_3	V_1	V_2	V_3
a. Freq of Recalibration	.35	.0	.50	1.00	0	.17	.35
b. Dependability	.35	.75	1.00	.34	.26	.35	.12
c. Ability to Cal Add Eq	.20	.96	.90	.80	.19	.18	.16
d. Diagnostic Capability	.10	.70	.80	.95	.07	.08	.10
			Totals		.52	.78	.73

3. Operator Opinion Survey

The opinion survey developed and administered by the Army Research Institute will be used to obtain the score on this factor. The survey questions are attached. The results of this survey will be converted into a single score on the interval [0, 1.0] for each vendor. The process for performing this transformation is described below.

For scoring purposes, the questionnaire addresses three issues. These are:

- The degree of satisfaction with each of the manual and automatic calibration systems.
- The degree of preference of each of the automatic calibration systems over the manual system.
- The adequacy of the training received for each of the automated calibration systems.

Two other elements of information, the number of training hours needed for each automated system and the rank order of preference for the manual and each of the automated systems, are also obtained. These elements cannot be used in directly converting the survey results into a score; however, the rank ordering of preference can be compared to the rank order obtained by scoring to check for consistency.

The degree of satisfaction with each of the calibration systems can be obtained from questions 1-8 and from question 15. For the purposes of evaluation, the value of the information content in the subset consisting of questions 1-7 is assumed to be the same as that in questions 8 and 15. (Admittedly, the information is partitioned in the first seven questions; however, the answer to question 1 clearly should not be given as much weight as the answer to question 15.) Assume that the answer to question 1 has the same degree of importance as that for questions 2, 3, . . . 7. Then the average of the scores for the first seven questions is a measure of the combination of all the issues addressed by those questions, and this measure, from the preceding discussion, is another estimate of the score for questions 8 and 15. Furthermore, the relative importance of these should be the same.

Before using the scores on the questionnaires to obtain a score for each vendor, the results will be examined for erratic opinions using appropriate statistical methods (ANOVA; non-parametric tests, confidence intervals, tests for outliers, etc.). Samples that do not belong to the population will be excluded. The score for each question will then be taken as the average score over all operators for that question. These operator averages will be designated as $S_{M,i}$ or $S_{v,i}$, where the subscript M refers to the manual system, the v refers to the AMCS of vendor v, and i refers to the question number. Using this convention, the score for the first seven questions would be

$$\bar{S}_M = \frac{1}{7} \sum_{i=1}^{7} S_{M,i}$$

and

$$\bar{S}_v = \frac{1}{7} \sum_{i=1}^{7} S_{v,i} .$$

The information content for $\bar{S}_M$ would be equivalent to that for $S_{M,8}$ and $S_{M,15}$, while that for $\bar{S}_v$ would be equivalent to $S_{v,8}$ and $S_{v,15}$.

The degree of preference for each vendor's AMCS can also be obtained from questions 1-7 by taking the difference between the score given to the manual system and that given to vendor v's AMCS. This leads to

$$\Delta \bar{S}_v = \bar{S}_v - \bar{S}_M.$$

As with the degree of satisfaction, the information content for $\Delta \bar{S}_v$ would be equivalent to that in $S_{v,13}$.

The range of possible values for scores on question 13 is from +100 to -100, with +100 corresponding to vendor v's AMCS being "very much better" than the manual system, 0 corresponding to their being equal, and -100 corresponding to its being "very much worse". Values less than 0 represent instances where the manual system is preferred to the automated system. The applicable range of scores on question 13 will be converted to a score on the interval [0-1.0] by adding 100 and dividing by 200. A similar argument applies to the transformation of the $\bar{S}_v$, $S_{v,8}$, $S_{v,9}$, and $S_{v,15}$ scores.

The range of possible values for the ΔS_v scores is +200 to -200. Hence, this score will be transformed to the interval [0-1.0] by adding 200 and dividing by 400.

On question 10, an opinion score of 0 indicates the desired condition. Opinion scores at either extreme are undesirable, but too much training is preferable to too little. Therefore, the following transformation will be used:

Survey Answer:	Way too Little	Too Little	Way too Much	Too Much	About Right
$S_{v,10}$	0	.25	.50	.75	1.00

In order to combine the scores on the questionnaire, it is assumed that all terms having a similar score on a question have the same utility. E.g., the result "somewhat satisfactory" on question 1 has the same utility as the term "somewhat convenient" on question 6, "somewhat easy" on question 7, and "somewhat better" on question 13.

The overall score for the user assessment factor will be the weighted sum of the scores on the interval [0-1.0] using the following weights:

Question Score	Weight
$\bar{S}_v$	.05
$S_{v,8}$	.05
$S_{v,15}$	.05
ΔS_v	.075
$S_{v,13}$	.075
$S_{v,9}$	.35
$S_{v,10}$	.35

B. Combination of Factor Scores.

The process of selecting the preferred vendor will make use of the additive weighted multi-attribute model, with each of the three factors constituting one attribute. The weights for each of the three attributes will be k_1, k_2, and k_3. In order to combine the three factors into one weighted score, all scores must be on the same scale. The Equipment Performance and Operator Opinion Survey scores are already on a scale of [0,1.0]. To transform the potential cost savings into the same scale, the Total Savings for each of the assignment alternatives will be divided by the maximum cost savings. In the event of a negative Total Savings for any alternative, each of the Total Savings will be divided by the range of Total Savings (Max-Min), and the score of 0 will be assigned to the minimum Total Savings alternative. These total cost savings, in the example given above, are as follows:

Alternative	Total Savings V_1	V_2	V_3	V_1V_2	V_1V_3	V_2V_3
1. Award basic buy of 5 to one vendor.	83	53	51			
2. Award basic buy of 10 to one vendor.	105	67	68			
3. Multiple award of 5 to each of 2 vendors.				122	118	91

The maximum savings is 122. Normalizing by this value gives the results in Table V.

Table V. Normalized Cost Savings Scores

Alternative	Savings V_1	V_2	V_3	V_1V_2	V_1V_3	V_2V_3
1. Award basic buy of 5 to one vendor.	.68	.43	.41			
2. Award basic buy of 10 to one vendor.	.86	.55	.56			
3. Multiple award of 5 to each of 2 vendors.				1.00	.97	.75

To permit the incorporation of the Equipment Performance and Operator Opinion Survey scores for the Multiple Award alternative, these scores will be the average of the scores on the factor for the two vendors receiving the multiple award. Thus, using the values in Table IV, the score for Equipment Performance for the V_1V_2 alternative would be $(.52 + .73)/2 = .625$. Similar calculations would be carried out for the remaining alternatives and the Operator Opinion Survey.

For the final calculations, assume $k_1 = .7$, $k_2 = .2$, $k_3 = .1$, and that the scores for the Operator Opinion Survey are .68, .71, and .81 for vendor's V_1, V_2, and V_3.* The total weighted score for each alternative is obtained as illustrated below. For the award of a basic buy of 5 to one vendor

$$S_{TV1} = .7\ (.68) + .2(.52) + .1\ (.68) = .648$$

$$S_{TV2} = .7\ (.43) + .2(.78) + .1\ (.71) = .528$$

$$S_{TV3} = .7\ (.41) + .2(.73) + .1\ (.81) = .514$$

*NOTE: Actual weights (values of k_1, k_2, k_3 are safeguarded by the contracting officer.

For the multiple award case,

S_{TV1V2} = .7(1.00) + .2(.52 + .78)/2 + .1 (.68+ .71)/2 = .8995

S_{TV1V3} = .7(.97) + .2(.52 + .73)/2 + .1 (.68+ .81)/2 = .8785

S_{TV2V3} = .7(.75) + .2(.78 + .73)/2 + .1 (.71+ .81)/2 = .7520

Similar results would be obtained for the award of a basic buy of 10 to a single vendor. The preferred approach is to award the contract to the vendor (or pair of vendors) which has the highest total weighted score. For the example, this would be a multiple award to vendors V_1 and V_2.

GROUND RULES FOR COST ADVANTAGE CALCULATIONS

1. All cost elements will be validated by MICOM Comptroller.

2. MICOM Comptroller provided discount rates will be used.

3. Times from the test phase used in the cost advantage model will be averaged over operators and over the number of times a meter underwent a particular calibration operation.

4. The total time it took each of the 18 operators to complete the manual tests will be subjected to an appropriate nonparametric test to determine if there are any significant differences in the mean performance of the operators. Any operators showing a significant difference in mean calibration time at the 10% level of significance will be eliminated from consideration as an operator of the AMCS.

5. All of the operators not eliminated in 4. above will be included in the determination of manual operating times. From this same group, 9 operators will be selected at random to test the automated calibration equipment.

6. Credit will be given to all vendors for a capability to adjust meters more rapidly than can be done with the manual system. This credit will be constant for each vendor at each site.

7. To account for the fact that the experimentally obtained calibration times include only productive activities, all experimental calibration times will be multiplied by 1/.66.

8. The lowest wage rate available at a fixed site will be used for the autocal systems provided it is not less than a GS-5. The average wage rate at a site will be used for the manual system.

AMCS TEST QUESTIONNAIRE #4

Please provide ratings of your Cycle 4 Automated Meter Calibrator at the end of Test Cycle 4.

Operator's Name ______________________________ Today's Date ______________

Calibrator used in Cycle 4 (Check one): ( ) Fluke ( ) Rotek ( ) Valhalla

1. How satisfactory or unsatisfactory is the Cycle 4 Calibrator for performance checking the meters' DC Voltage function?

Completely Satisfactory	Mostly Satisfactory	Somewhat Satisfactory	Slightly Satisfactory	Borderline	Slightly Unsatisfactory	Somewhat Unsatisfactory	Mostly Unsatisfactory	Completely Unsatisfactory
( )	( )	( )	( )	( )	( )	( )	( )	( )
+100	+75	+50	+25	0	-25	-50	-75	-100

2. How satisfactory or unsatisfactory is the Cycle 4 Calibrator for performance checking the meters' AC Voltage function?

Completely Satisfactory	Mostly Satisfactory	Somewhat Satisfactory	Slightly Satisfactory	Borderline	Slightly Unsatisfactory	Somewhat Unsatisfactory	Mostly Unsatisfactory	Completely Unsatisfactory
( )	( )	( )	( )	( )	( )	( )	( )	( )
+100	+75	+50	+25	0	-25	-50	-75	-100

3. How satisfactory or unsatisfactory is the Cycle 4 Calibrator for performance checking the meters' DC Current function?

Completely Satisfactory	Mostly Satisfactory	Somewhat Satisfactory	Slightly Satisfactory	Borderline	Slightly Unsatisfactory	Somewhat Unsatisfactory	Mostly Unsatisfactory	Completely Unsatisfactory
( )	( )	( )	( )	( )	( )	( )	( )	( )
+100	+75	+50	+25	0	-25	-50	-75	-100

4. How satisfactory or unsatisfactory is the Cycle 4 Calibrator for performance checking the meters' Resistance function?

Completely Satisfactory	Mostly Satisfactory	Somewhat Satisfactory	Slightly Satisfactory	Borderline	Slightly Unsatisfactory	Somewhat Unsatisfactory	Mostly Unsatisfactory	Completely Unsatisfactory
( )	( )	( )	( )	( )	( )	( )	( )	( )
+100	+75	+50	+25	0	-25	-50	-75	-100

5. How satisfactory or unsatisfactory are the visual displays (screen and dials) of the Cycle 4 Calibrator?

Completely Satisfactory	Mostly Satisfactory	Somewhat Satisfactory	Slightly Satisfactory	Borderline	Slightly Unsatisfactory	Somewhat Unsatisfactory	Mostly Unsatisfactory	Completely Unsatisfactory
( )	( )	( )	( )	( )	( )	( )	( )	( )
+100	+75	+50	+25	0	-25	-50	-75	-100

6. How convenient or inconvenient is it to use the knobs and switches of the Cycle 4 Calibrator?

Extremely Convenient	Very Convenient	Somewhat Convenient	Slightly Convenient	Borderline	Slightly Inconvenient	Somewhat Inconvenient	Very Inconvenient	Extremely Inconvenient
( )	( )	( )	( )	( )	( )	( )	( )	( )
+100	+75	+50	+25	0	-25	-50	-75	-100

7. How easy or difficult is it to set up most meters for calibration with the Cycle 4 Calibrator?

Extremely Easy	Very Easy	Somewhat Easy	Slightly Easy	Borderline	Slightly Difficult	Somewhat Difficult	Very Difficult	Extremely Difficult
( )	( )	( )	( )	( )	( )	( )	( )	( )
+100	+75	+50	+25	0	-25	-50	-75	-100

8. Overall, how satisfactory or unsatisfactory is the Cycle 4 Calibrator for performance checking the various meter functions?

Completely Satisfactory	Mostly Satisfactory	Somewhat Satisfactory	Slightly Satisfactory	Borderline	Slightly Unsatisfactory	Somewhat Unsatisfactory	Mostly Unsatisfactory	Completely Unsatisfactor
( )	( )	( )	( )	( )	( )	( )	( )	( )
+100	+75	+50	+25	0	-25	-50	-75	-100

9. Rate the quality of the training/instruction you received in Cycle 4.

Completely Satisfactory	Mostly Satisfactory	Somewhat Satisfactory	Slightly Satisfactory	Borderline	Slightly Unsatisfactory	Somewhat Unsatisfactory	Mostly Unsatisfactory	Completely Unsatisfacto
( )	( )	( )	( )	( )	( )	( )	( )	( )
+100	+75	+50	+25	0	-25	-50	-75	-100

10. Overall, the amount of training time taken for Cycle 4 training was:

Way Too Much	Too Much	About Right	Too Little	Way Too Little
( )	( )	( )	( )	( )
+100	+50	0	-50	-100

11. How many hours of training did you receive in Cycle 4? ________ Hours

12. Please estimate the number of training hours you believe should have been given in Cyc 4: ________ Hours