Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Erik D. Fosshage

Entitled
The Effect of Job Performance Aids on Quality Assurance

For the degree of Master of Science in Industrial Engineering

Is approved by the final examining committee:

Barrett Caldwell
Chair

Caren Wenner

Mark Lehto

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Barrett Caldwell

Approved by: Abhi Deshmukh 4/17/2014
Head of the Graduate Program Date

THE EFFECT OF JOB PERFORMANCE AIDS ON QUALITY ASSURANCE

A Thesis

Submitted to the Faculty

of

Purdue University

by

Erik Fosshage

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Industrial Engineering

May 2014

Purdue University

West Lafayette, Indiana

UMI Number: 1564868

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI 1564868

Published by ProQuest LLC (2014). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ACKNOWLEDGEMENTS

This master's thesis is a collaboration between the Surety Assessment, Engineering, and Analysis Center at Sandia National Laboratories (SNL) and the School of Industrial Engineering at Purdue University. Funding was provided by the University Part Time program at SNL as well as the Nuclear Safety Research and Development Program within the National Nuclear Security Administration of the United States Department of Energy.

The author wishes to thank Professor Barrett Caldwell at Purdue University and Caren Wenner at Sandia National Laboratories for technical guidance and direction during this research. Additional advice at SNL was provided by Courtney Dornburg, Susan Stevens-Adams, Kerstan Cole, Judi See, Ed Thomas, Jeffrey West, Jennifer Samora, Jenny Lindblom, and Leslie Simmons. Administrative support was provided by Shelby Hopkins, Gwen Zon, Lori Hatcher, Julie Hillskemper, Laura Calderone, Missy Hess, Susan Fourt, Tiffany Vargas, Amy Rein, Rogulja Wolf, Gary Boulanger, Tom Wubbels, and Ann Osburn.

Lastly, the author has been honored to engage in a correspondence with Dr. Edgar L. Shriver; a true pioneer in the human factors community regarding task analysis and job performance aids.

TABLE OF CONTENTS

Page
LIST OF TABLES iv
LIST OF FIGURES v
ABSTRACT vi
CHAPTER 1. INTRODUCTION 1
1.1 Motivation 2
1.2 Research Statement 4
1.3 Thesis Organization 4
CHAPTER 2. LITERATURE REVIEW 6
CHAPTER 3. METHODOLOGY 13
3.1 Experimental Task Selection 13
3.2 Test Approach 17
3.3 Sample Size Determination 22
3.4 Test Procedure 23
3.5 Job Performance Aid 29
CHAPTER 4. RESULTS AND ANALYSIS 32
CHAPTER 5. DISCUSSION 39
CHAPTER 6. CONCLUSION AND FUTURE RESEARCH 43
REFERENCES 47
APPENDICES
Appendix A Pattern A 54
Appendix B Pattern B 56
Appendix C Incorrect Parts 58
Appendix D Job Performance Aid 59
Appendix E Raw Data 60
Appendix F Fault Types 64
Appendix G Minitab Results 72
Appendix H Assembly Instructions from Pattern A (Lego™ Kit 5508) 83
Appendix I Assembly Instructions from Pattern B (Lego™ Kit 6167) 103

LIST OF TABLES

Table Page

Table 1: Definitions Used in the Assembly Task 18
Table 2: Probability of Correctly Concluding that p2>p1 23
Table 3: Grouping of Experimental Participants By Sequence of Assembly 25
Table 4: Fault Descriptions/Types 26
Table 5: Fault Detection Performance, By Participant 32
Table 6: Performance, By Fault Number 33
Table 7: Logistic Regression Table, Pattern A 35
Table 8: Logistic Regression Table, Pattern B 36
Table 9: Estimated Probability of Detection, Pattern A 37
Table 10: Estimated Probability of Detection, Pattern B 38

LIST OF FIGURES

Figures Page

Figure 1: Test Setup 24
Figure 2: Sample Instructions Page, Depicting Fault #1 (Fault Type 1) 27
Figure 3: Fault #1 Occurring in an Experimental Trial 28

Appendix Figure

Figure A1: Pieces from Pattern A (Lego™ Kit 5508) 54
Figure A2: Lego™ Kit 5508 Contents 55
Figure B1: Pieces from Pattern B (Lego™ Kit 6167) 56
Figure B2: Lego™ Kit 6167 Contents 57
Figure C1: Parts Used for Faults 58
Figure E1: First Pattern Raw Data (1 of 2) 60
Figure E2: First Pattern Raw Data (2 of 2) 61
Figure E3: Second Pattern Raw Data (1 of 2) 62
Figure E4: Second Pattern Raw Data (2 of 2) 63
Figure F1: Sample Instructions Page, Depicting Fault #1 (Fault Type 1) 64
Figure F2: Fault #1 Occurring in an Experimental Trial 65
Figure F3: Sample Instructions Page, Depicting Fault #8 (Fault Type 2) 66
Figure F4: Fault #8 Occurring in an Experimental Trial 67
Figure F5: Sample Instructions Page, Depicting Fault #3 (Fault Type 3) 68
Figure F6: Fault #3 Occurring in an Experimental Trial 69
Figure F7: Sample Instructions Page, Depicting Fault #5 (Fault Type 4) 70
Figure F8: Fault #5 Occurring in an Experimental Trial 71

ABSTRACT

Fosshage, Erik D. MSIE, Purdue University, May 2014. The Effect of Job Performance Aids on Quality Assurance. Major Professor: Barrett Caldwell.

Job performance aids (JPAs) have been studied for many decades in a variety of disciplines and for many different types of tasks, yet this is the first known research experiment using JPAs in a quality assurance (QA) context. The objective of this thesis was to assess whether a JPA has an effect on the performance of a QA observer performing the concurrent dual verification technique for a basic assembly task. The JPA used in this study was a simple checklist, and the design borrows heavily from prior research on task analysis and other human factors principles. The assembly task and QA construct of concurrent dual verification are consistent with those of a high consequence manufacturing environment. Results showed that the JPA had only a limited effect on QA performance in the context of this experiment. However, there were three important and unexpected findings that may draw interest from a variety of practitioners. First, a novel testing methodology sensitive enough to measure the effects of a JPA on performance was created. Second, the discovery that there are different probabilities of detection for different types of error in a QA context may be the most far-reaching results. Third, these results highlight the limitations of concurrent dual verification as a control against

defects. It is hoped that both the methodology and results of this study are an effective baseline from which to launch future research activities.

CHAPTER 1. INTRODUCTION

According to James Cantrell, the main engineer for the Skipper satellite, “It’s always the simple stuff that kills you.” Skipper failed one day into its mission because Russian scientists mistakenly connected the solar panels backwards (“Russians Miswire Satellite”, 1996), an oversight that could have been avoided with the use of a quality assurance (QA) checklist. A checklist may also have prevented the NASA Genesis probe from crash-landing in 2004 when its drogue parachute failed to deploy due to a deceleration sensor that was installed backwards (“Official: Genesis Pre-Launch”, 2006). These examples remind us that even simple human errors can have large consequences, and though human error can never completely be eliminated we can nonetheless introduce processes and tools that are designed to prevent errors or to catch errors and minimize their impact.

One purpose of checklists is to support either simple or complex actions that are subject to the limitations of human performance. In the delivery of a product or activity, an overarching goal for quality assurance and human factors professionals is often the prevention of defects, and this becomes equally as important as it is difficult when dealing with both complex and critical systems. People who are present during high consequence failures (or near misses) always remember the story and are extremely

unlikely to repeat the same mistakes, a concept best captured by Vilfredo Pareto in 1870: "Give me a fruitful error any time, full of seeds, bursting with its own corrections" (from *Comment on Kepler*, quoted in Helmreich and Merritt, 2001). A checklist often is a lesson intended for an audience that was not present during a mishap who should learn the failure mechanisms in order to avoid them. This study examines how a checklist (a specific type of job performance aid, or JPA) may affect human performance in the verification activities that are often employed in systems that have high consequences for failure.

1.1 Motivation

The author's original intent in selecting a thesis topic was to bridge the QA and human factors disciplines, because his work experience resides in the former and major field of study is in the latter. Consultations with staff in the Human Factors department at Sandia National Laboratories (SNL) revealed an internal paper titled "The Effectiveness of Verification in Industrial Applications" (Stevens-Adams et al., 2013) that reviews the academic and applied literature on the verification construct, variables underlying its effectiveness, and the applications that are best suited to a particular work environment. The paper cites different definitions of verification from Department of Energy (DOE) standards and handbooks, including the concurrent dual verification technique that is utilized by the Surety Engineering Group[1] at SNL in its high consequence manufacturing environments (DOE, 1993; DOE, 2009) whereby a QA observer is present to witness

[1] "Surety" encompasses a number of disciplines related to nuclear weapons such as safety, reliability, quality, human factors, and assessment, and is a term not widely used outside of SNL. The Surety Engineering Group is staffed primarily by quality engineers.

activities being performed. In consultation with SNL and Purdue advisors, it was determined that there has been a dearth of research on the concurrent dual verification technique and that an appropriate thesis topic would be to address this particular QA task.

An episode from the author's past involved mentoring an inexperienced QA checker to verify assembly tasks for a high consequence product, including questions about how to specifically complete these tasks. This particular incident related to the performance of concurrent dual verification led to the selection of a job performance aid as the focus of this thesis, and more specifically an examination of checklist aids in QA. After performing a cursory literature search on JPAs, it was initially determined that the topic of JPAs related to QA tasks had not previously been studied. A more thorough search of the JPA literature eventually located a detailed experiment (Shriver et al., 1982) that studied the effect of JPAs on eight (8) different tasks within a nuclear power plant, one of which was specific to QA. However, the QA-focused JPA was eventually dropped from the Shriver study as being too task-interdependent and reliant upon wider functional understanding from an expert operator. The project director (L. Zach, personal communication, June 2013) and lead principal investigator (E.L. Shriver, personal communication, June 2013) of the 1982 report have been located and verified that, to their knowledge, no other research on JPAs for QA has ever been published.

This research study is therefore a nexus for multiple interests: the design and use of checklists to improve human performance, the linkage of human factors and QA

techniques for defect prevention, and a continuation of both previous and existing research. In addition to academic discoveries, this study of JPAs can benefit a wide-ranging audience of practitioners as a tool for reducing the likelihood of high consequence events that still happen in every-day life; rockets and satellites continue to fail (Cheng et al., 2009), medical errors lead to tragic results (Dhillon, 2008), and in fact lives could even be saved by implementing a checklist (Haynes et al., 2009). This topic has such an impact that it has even spawned a New York Times bestseller *The Checklist Manifesto* (Gawande, 2010) that has drawn popular interest on an international level.

1.2 Research Statement

Since this was the first known study that examined the use of JPAs for the QA role, the primary research focus was to determine whether a simple JPA (i.e., checklist) has an effect on QA performance, and then only on the concurrent dual verification technique. Once the experimental design was chosen, additional hypotheses were tested regarding independent variables and an important test assumption that was inadvertently germane to the results.

1.3 Thesis Organization

This thesis consists of six chapters. Chapter 2 examines prior research on JPAs in a variety of disciplines, especially human factors, and provides definitions of concurrent dual verification in a QA context. Chapter 3 explains the methodologies used in the design of the experiment, as well as the JPA itself. Chapter 4 presents results and

analysis of the data while Chapter 5 discusses the key implications of the study. Chapter 6 presents conclusions and recommends future studies regarding the effect of different JPA formats on QA performance, as well as their potential to affect other QA tools and processes besides verification techniques..

CHAPTER 2. LITERATURE REVIEW

Probably the best-known JPA is the pilot's checklist, which was born on October 30, 1935 at Wright Air Field in Dayton, Ohio during the initial test flight of the Boeing Model 299 aircraft (which later became the B-17 "Flying Fortress"). The Model 299 crashed after take-off, killing two of the 5 crewmen. The accident investigation board determined that no mechanical failures had occurred and that the aircraft crashed because the pilot forgot to unlock the elevator and rudder locks. The Model 299 was substantially more complex than previous aircraft and simply too much for one pilot to remember how to fly, so the approach at the time was to simplify this complexity with a pilot's checklist (Meilinger, 2004; Gawande, 2010).

Over the next few decades research on JPAs would be conducted in a variety of disciplines, beginning with U.S. Air Force behavioral research in the 1950s. From 1951 to 1956 the Air Force Behavioral Science Laboratory, at the re-named Wright-Patterson Air Force Base in Dayton, contracted the human factors specialists at the American Institute for Research to study how behavioral and psychological factors could be identified and then used to design training requirements for specific tasks (Miller, 1953). It was during this study that Dr. Edgar Shriver coined the term "task analysis" to specify the need to identify stimulus elements, or behavioral cues, that indicate when a task is

complete and then place these directly into training instructions (Shriver et al., 1982). Newman (1957) and Miller (1956), under contract to the Air Force Personnel and Training Research Center in Denver, Colorado, also recognized the need to identify specific behavioral processes to improve both accuracy and training time and then elicit these processes into clear and concise instructions that are compatible with the skills and abilities of the users.

Task analysis was widely adopted and matured by a variety of practitioners, but behavioral cues and the role of behavioral processes were not always used due to inconsistent applications of the technique as new task analysis methodologies were developed (Davies, 1973; Stammers et al., 1990). For example, Mager (1962) renamed the stimulus elements as "conditions" and this resulted in the stimulus term often being misinterpreted as environmental (e.g. temperature) or setup conditions to the experiment; as a result the *behavioral* stimulus elements were often lost (Shriver et al., 1982).

In 1969 the Special Interest Group on Task Analysis of the Human Factors Society (chaired by Shriver) counted over 80 variations of task analysis in use at the time; in many instances the behavioral cues were missing. This led to the sub-categorization of task analysis methods into more specific methodologies such as basic task analysis (Miller, 1953), job analysis (Chapanis, 1970), functional analysis (Shriver, 1960), and link analysis (Thomson, 1972). A major output of the interest group's effort was the establishment of the behavioral task analysis (BTA) technique that returned the stimulus element to task analysis. The BTA technique was later codified in a U.S. Air Force

technical report (Shriver, 1975) that would also serve as an important milestone in the development of JPAs.

The first research activities that specifically used the term "job performance aid" began with more U.S. military contracts in the late 1950s and early 1960s in conjunction with task analysis (e.g. Shriver, Fink, & Trexler, 1959; Folley, 1961; Folley & Shettel, 1962; Goff et al., 1969). Over the next two decades JPAs were being studied in a variety of contexts. The early military studies tended to look for ways to reduce errors in complex tasks that were not often performed correctly (such as repair and maintenance activities) or to shorten the training time necessary for novice technicians to acquire such skills (Gebhard, 1970; Elliott & Joyce, 1971; Foley, 1972). Later studies were refined to determine how different JPA formats (pictures or printed text) could convey information differently and were therefore better suited for particular circumstances (Booher, 1975; Smillie, 1978; Smillie & Ayoub, 1980). Throughout these experiments the JPA was found to be an effective tool at simplifying tasks for novice users that would normally require extensive training or complex information processing, and errors in human performance were reduced significantly. These experiences led to a variety of specifications (Folley et al., 1971; Joyce et al., 1973a; Shriver, 1975) and handbooks (Joyce et al., 1973b; Booher, 1978; Smillie, 1985) to assist developers in creating effective JPAs for a wide variety of tasks in both military and commercial industries.

Meanwhile, specific industries were conducting their own studies on the limitations of human performance. McKenzie (1958) was an early proponent of having clear standards

and instructions in order to improve the accuracy and consistency of inspectors. This followed with more than three decades of research from the industrial inspection community (summarized in detail by Fox, 1973; Wiener, 1975, Drury, 1982, and See, 2012), with some of it focused on vigilance theory (e.g. Elliott, 1960; McGrath and Harabedian, 1961; Mackie, 1964; Baddeley and Colquhoun, 1969, Tsao, Drury, and Morawski, 1979; Tsao and Wang, 1984; Murray and Caldwell, 1996) that stemmed from the pioneering work of Mackworth (1950). These studies were paralleled by researchers who began to understand the underlying models of human information processing that would become useful to JPA developers (Gagne, 1962; Kibler, 1965; Harris, 1969, Norman, 1981; Rasmussen, 1982; Norman, 1983; Rasmussen and Vicente, 1989; Reason, 1990; Barshi and Healy, 1993; Allen and Rankin, 1997).

Following the Three Mile Island accident (Kemeny, 1979) the nuclear power industry commissioned a series of studies to understand and eliminate the potential for human errors, and JPA research experienced a resurgence of interest (Clark, 1982; Shriver et al., 1982; Mattson, 1989; Hallbert et al., 1992). In the years since, other high consequence industries such as commercial aviation and medicine have followed suit by embracing a JPA focused approach (most notably checklists) as well as wider principles of error management, human factors, and quality assurance (Helmreich and Schaefer, 1994; Helmreich, 2000; Helmreich and Merritt, 2001; Haynes et al., 2009; Gawande, 2010; White et al., 2010).

SNL designs and builds a variety of high consequence products, however its primary mission is with nuclear weapons. SNL is therefore subject to the guidance and oversight by the Department of Energy (DOE). As reviewed in Stevens-Adams (2013), the DOE *Human Performance Handbook* (DOE, 2009) focuses on tools and methods for preventing and mitigating human error, including processes specifically related to verification activities that are useful for quality assurance:

- **Self-Checking** – A process by which the performer pauses to reflect on the intended action and its expected outcome, whether the proposed action is correct for the situation, and resolve any questions or concerns *before* proceeding by using the STAR (Stop, Think, Act, Review) practice.
- **Peer Checking** – A series of actions by 2 individuals working together at the same time and place, before and during a specific action, as an *augmentation but not a replacement* for self-checking.
- **Concurrent Verification** – A series of actions by 2 individuals, at the same time and place, to separately confirm the condition before, during, and after an action, when the consequences of an incorrect condition would subsequently lead to undesired harm.
- **Independent Verification** – A process by which one individual, separated by time and distance from the action, confirms the desired condition.
- **Peer Review** – A process by which qualified reviewers read and check the quality of another's work, and is an informal technique that does not supplant required procedures or checklists.

As further discussed in Stevens-Adams (2013), DOE Order 422.1, *Conduct of Operations*, requires independent and concurrent dual verification techniques to be implemented for critical equipment configurations. These two techniques are further defined in the DOE *Guide to Good Practices for Independent Verification* (DOE, 1993):

- **Independent Verification** – The act of checking, by a separate qualified person, that a given operation or the position of a component conforms to established criteria.
- **Concurrent Dual Verification** – A method of checking an operation, an act of positioning, or a calculation in which the verifier independently observes and/or confirms the activity.

There has been a divergence between the term *inspector*, which implies an independent reviewer of a particular item or task, and *checker* that is specified by the verification techniques. As a result of the above definitions (plus the fact that this study is about *check*-lists), the term *checker* will be used throughout this paper to refer to the QA role being performed by the experimental participants.

The motivation for this study should now be understood as an intersection between the complementary disciplines of human factors and QA (itself derived from the tradition of inspection), with a focus on the concurrent dual verification activity within high consequence environments. The JPA approach continues a contemporary interest in checklists while leveraging the legacy techniques of task analysis, plus it is one of many tools that a QA practitioner can use as a control against the risk of defects. The

concurrent dual verification JPA thus fills a research void and should draw interest from a variety of disciplines.

CHAPTER 3. METHODOLOGY

This chapter describes the methodology and approach used for the research study. Section 3.1 specifies the approach behind the selection of the experimental task and presents an appropriate and novel test apparatus as a contribution to the wider body of knowledge. Section 3.2 explains the experimental design. Section 3.3 explains how the number of experimental participants was determined and Section 3.4 discusses the selection of participants and the presentation of the experiment. Section 3.5 details the rationale and configuration of the job performance aid used in this study.

3.1 Experimental Task Selection

The primary emphasis of the original research question is on the effectiveness of a JPA on a concurrent dual verification task. However, there must be a task for the QA checkers to observe such that they can perform concurrent dual verification. Basic guidelines for selecting this task were as follows:

- Not be too complex or time-consuming
- Not be too simple, such that the ability to inject faults would be difficult as they would be too obvious

- Be consistent with a task that may be conducted in a high consequence manufacturing environment

A predefined Lego™ assembly task (with visual assembly instructions) was determined to satisfy these conditions, with a corresponding checklist for the QA checker to follow (or have absent in the control condition). An advantage of selecting a Lego™ assembly task over more applied techniques (such as repairing a lawnmower engine or replacing a circuit board in a personal computer) is that there is not any built-in covariance of prior knowledge for experimental participants who may have expertise in these applied tasks. In other words, selecting simple tasks would reduce uncontrolled differences in participant expertise. Another advantage is the similarity of the Lego™ task with those conducted in a high consequence manufacturing environment. Assemblers are often provided with a kit of similar-looking parts (e.g., fasteners of different size but equal length) and instructions how to install them, and in many real-world applications there is a QA checker to oversee the task and ensure that no errors are made.

A within-subjects design model was chosen in order for participants to have their performance evaluated both with and without the checklist, so this drove the need for a second assembly task. Due to the potential for learning effects, two *different* Lego™ assembly tasks (Pattern A, Pattern B) were identified in which faults could be introduced by the inclusion of incorrect pieces into the patterns. Another variable to be reasonably controlled is the difference in size or complexity between the two patterns. This constraint eliminated many of the popular Lego™ patterns from contention, such as

Lego™ City or themed sets like Star Wars™ or super heroes, since the ability to equalize both size and complexity across two patterns is difficult. Further, defined patterns with uniquely shaped pieces typically require all parts to be used during assembly (i.e., there are no pieces left over). This might make it too obvious to inject faults into these complex patterns. The option chosen for this experimental task was to select two similar but existing patterns (so the instructions were already developed and available) and then combine them into a larger boxed kit (thus each pattern was a sub-assembly within the larger kit).

Pattern A (see Appendix A) has 104 pieces with 7 of them specifically chosen for inserting a fault. Pattern B (see Appendix B) has 150 pieces, with another 7 pieces specifically chosen for inserting a fault. Since one of the constraints is to avoid selecting a task that is too simple, the patterns were not assembled as stand-alone kits but instead contained within a larger set of pieces that may or may not be used in either sub-assembly. This eliminates the potential confound where the QA checker might notice that an "incorrect" part was used if there are supposed to be no pieces remaining when the assembly task is completed. The total number of available pieces needs to be higher than both sub-assemblies combined, with additional margin such that there is sufficient uncertainty in the experiment to warrant the use of a JPA. Since the total number of pieces in Pattern A and Pattern B is 254 (almost $256 = 2^7$ or 7 bits of information), a total assembly kit of 512 pieces was chosen since $512 = 2^9$ or 9 bits of potential information (Posner, 1964).

However, the total number of bits of information is much higher than 9 bits when considering the different part shapes (23 for Pattern A), colors (6 for Pattern A), and markings that each act as multipliers to the total number of possible combinations of parts within the two patterns. This large number is roughly doubled when adding in Pattern B, and doubled again when considering parts not used in either pattern. Thus there is tremendous (and sufficient) uncertainty in the kit of 512 parts that experimental participants were not able to determine the total number of pieces being used for each pattern. If the JPA effectively improves the chances that the *correct piece* in the *correct color* and *correct shape* is installed in the *correct way*, and it is "infused" (Oberhettinger, 2012) at the *correct time* and *correct place* in the assembly process, then it would be a tremendously useful tool to support concurrent dual verification in mitigating human error.

In order to more fully describe the correct piece / correct place discussion above, consider that each of the 512 pieces belongs to only one of the following five part groups:

- Parts used only in Pattern A: 52 pieces
- Parts used only in Pattern B: 98 pieces
- Parts that exist in both patterns A and B: 104 pieces (or 52 in both patterns)
- "Incorrect" parts that are used to interject faults into both patterns: 9 pieces (see Appendix C)
- Parts not used in either pattern: 249 pieces

To further clarify, Pattern A has 52 pieces that are unique and 52 that are also used in Pattern B, for a total of 104 pieces. Pattern B has 98 pieces that are unique and 52 that are also used in Pattern A, for a total of 150 pieces. Mutual exclusivity of the five part groups allows exact probabilities to be determined for analysis and compared to experimental results. The inclusion of both patterns in the kit of 512 pieces, as well as the decision to not return the first pattern's pieces to the kit before commencing with the second pattern in the trial, are counterbalanced by the randomized assignment of participants to one of the four test conditions.

3.2 Test Approach

In the Lego™ assembly task described in Section 3.1, experimental participants performed the role of QA checker and the author acted as the assembler of the patterns. A distinction was made between an ***error*** by the assembler that was purposely made (***fault***) in order to elicit experimental data and an error that was unintentionally made (***mistake***) and is less germane to the results. However, the use of the term "error" was necessary in providing instructions to the checkers because they correctly identified both faults and mistakes during the experiment without any knowledge of the difference between the two. The instructions given to the participant before the test began were: "Please let me know if you notice any errors. " When a fault was made and the QA checker did not notice, this was defined as a ***miss***; if they did notice, it was defined as a ***detection***. Some faults were detected immediately, but others were detected later in the assembly process. This difference was anticipated and considered important to the results, so there needed

to be a way to distinguish between the two responses. While the identification of a fault before turning the page in the assembly instructions is a detection, if the fault was identified after the page was turned it was determined to be a ***catch***. A ***false alarm*** occurred when an error (fault or mistake) was not performed but the QA checker notified the assembler that they thought one had occurred. If the assembler intentionally disguised a fault from the QA checker then it was classified as a ***violation*** (though the experiment was designed to eliminate the risk of violations and none were present in the results). Finally, ***sequence errors*** are a special type of false alarm that is similar to Norman's (1981) mode error, where the response is incorrect and represents a user assuming a state other than the true state of the task. These definitions, specific to this study and not necessarily in common use elsewhere, are summarized in Table 1:

Table 1: Definitions Used in the Assembly Task

Fault	An intentional and specific error that is introduced by the assembler
Mistake	An unintentional error by the assembler
Error	The use of an incorrect piece or incorrect assembly order
Miss	An error that is not noticed by the QA checker; includes catches
Detection	Before turning the next page of the assembly instructions, the QA checker identifies an error
Catch	An error that is noticed by the QA checker later (after the page has been turned)
False alarm	Any response from the QA checker when there is no error present
Violation	A fault that is purposely hidden from the QA checker, such that there is not an opportunity to notice it
Sequence error	The QA checker verbally authorizes the assembler to turn the page of the instructions before the assembly task on that page has been completed; treated as a false alarm

Faults were inserted by the assembler into both patterns in specific pre-determined locations, with the pre-test instruction that the QA checker observe the assembly task and tell the assembler if they noticed any errors. The fault types are:

1. *Marking Fault*: Insert the incorrect piece that has markings. For example, instead of inserting the "correct" 2 x 2 yellow piece another 2 x 2 yellow piece with a pattern on one side is installed.
2. *Incorrect Piece:* Insert the incorrect piece(s) but with no markings. For example, instead of inserting the "correct" 2 x 4 black piece two 1 x 4 black pieces is installed.
3. *Wrong Order:* Insert the correct piece(s) but in the incorrect configuration. For example, instead of constructing a 2 x 10 wall with a 2 x 8 piece on the left and a 2 x 2 piece on the right, the order is switched and the 2 x 2 piece goes on the left
4. *Wrong Orientation:* Insert the correct piece(s) but in the wrong location or orientation. For example, a window piece is installed backwards, or offset by one row from what the "correct" orientation is.

The fault types were chosen for their reasonable similarity to those that occur in high consequence environments. For example, fault type 4 (wrong orientation) is similar to a component being installed backwards on a printed circuit board. The author has past experience with fault type 1, dating from an incident whereby a damaged part with noticeable indications of wear was nearly installed onto a larger assembly. The specific faults in the Lego™ assembly task were only selected for their ease of insertion and

ability to avoid detection by the QA checker, and were not equally spaced throughout the assembly task. These faults are further detailed in Section 3.4.

The independent variables are as follows:

- Pattern order (A first or B first)
- JPA (present or not present)

Uncontrolled variables (potential covariants) that were observed are:

- Pattern A elapsed time
- Pattern B elapsed time
- Patterns A and B combined elapsed time
- Fault type

Because the errors (both faults and mistakes) are performed by the assembler they are independent of the presence or absence of the JPA. Therefore the responses from the participant (the QA checker) represent the primary dependent variables for this study. For purposes of simplifying the statistical design the dependent variables have been categorized as described above and summarized into only 3 response levels: 1) detections; 2) misses, and 3) false alarms.

The primary purpose of this experiment is to determine whether or not the JPA has an effect on the performance of the QA checker. Therefore, the first null hypothesis to be tested is:

H_1: The presence of a JPA has no effect on the detection of faults in the quality assurance role.

with the alternative hypothesis being:

H_{1A}: The presence of a JPA ***has*** an effect on the detection of errors in the quality assurance role.

The following additional hypotheses are being tested in this experiment, specifically addressing the independent variables listed previously:

H_2: There is no difference in the detection of faults with a JPA than without a JPA.
H_{2A}: There **is** a difference in the detection of faults with a JPA than without a JPA.

H_3: The order of presentation of the JPA has no effect on the detection of faults in the quality assurance role.
H_{3A}: The order of presentation of the JPA **has** an effect on the detection of faults in the quality assurance role.

H_4: The order of presentation of the different patterns has no effect on the detection of faults in the quality assurance role.
H_{4A}: The order of presentation of the different patterns **has** an effect on the detection of faults in the quality assurance role.

An additional hypothesis, H_5, is explained in the next section.

3.3 Sample Size Determination

The minimum number of participants was determined from the estimated probability of detecting each fault, with different values estimated for with and without a JPA. The different probabilities of detection among the 14 total faults span a fuzzy range between somewhat-easy and somewhat-difficult, and the added complexity of 4 different fault types with uneven placement throughout the two patterns made it necessary to create simplifying assumptions for both easier statistical analyses and a testable hypothesis. The probabilistic Rasch (1960) model, which in turn requires an ordered index based on the Guttman (1944) scale, is an appropriate guide for this experiment as it models the trade-off between the QA checker's abilities and the difficulty of detecting each fault.

One simplifying assumption was to assign a constant average probability of detection across all fault instances. Since the checklist specifically identified the 4 different fault types, it was reasonable to assume that the average probability of detection is p_1=0.5 without a JPA and p_2=0.9 with a JPA. The number of experimental participants (n) must be a multiple of 4 to match the number of test conditions (listed in Section 3.1) and therefore make the experiment balanced. For n*14 binary trials (n*7 with a JPA and n*7 without a JPA), the probability of concluding that there is a difference when p_1=p_2 depends on n as follows in Table 2:

Table 2: Probability of Correctly Concluding that p2>p1

n	Type-1 error of 0.05	Type-1 error of 0.025
4	0.96	0.93
8	0.9995	0.998
12	~1	0.99998
≥16		~1

This analysis suggests a minimum sample size of eight participants. However, the simplifying assumption of average probability of detection across the individual 14 faults suggests that this number should be higher. Once the experimental trials began it soon became apparent that the addition of more participants was not particularly onerous on the assembler or support staff and could be accomplished in a reasonably short timeframe. Thus the number of participants for this experiment was raised to 24. In addition, the assumption of an average probability of detection being uniformly distributed across the experiment is an additional hypothesis to be tested:

H_5: The probability of detection for each fault is equal.

H_{5A}: The probability of detection for each fault is **not** equal.

3.4 Test Procedure

Participants consisted of permanent and contract employees (14 males, 10 females) from SNL. Candidates who perform QA activities as part of their normal job responsibilities were specifically excluded from this study, so all QA checkers were considered novices in this role. The experiment was conducted in a video recording studio to facilitate both

the capture of the assembly tasks and audio of the QA checker's interactions with the assembler. The video camera was placed such that all responses were recorded but not the faces of participants, in order to maintain confidentiality of participation. Figure 1 shows the test setup from one of the experimental trials with the assembler on the left and the QA checker seated off-camera on the right.

Figure 1: Test Setup

The assembly instructions were printed and used by the assembler in clear view of the QA checker, along with the collection of 512 pieces. These assembly instructions were the original instructions for each pattern developed by Lego™, with no changes or additions. The pre-test instructions given verbally to the QA checkers (only once, before the first pattern was presented) were to observe the assembly task and report any errors as soon as they were noticed, and not to touch any of the Lego™ pieces or assist the assembler in locating them. Participant #7 was disqualified due to misunderstanding the

pre-test instructions. Another participant was recruited to replace #7 prior to the completion of the study, maintaining a total of 24 valid participants.

For one of the two assembly tasks in each experimental trial, the checklist was provided for review beforehand and was available throughout the duration of that assembly task. Table 3 summarizes how the order of patterns varied across the 24 participants. By varying the order of assembly in this way, specific effects of the presence of a JPA (if they exist) can be distinguished from the effects of sequence of assembly ($H_2 - H_3$).

Table 3: Grouping of Experimental Participants By Sequence of Assembly

Number of Participants	Sequence of Assembly, Presence of JPA	Abbreviation
6	Pattern A without JPA, followed by Pattern B with JPA	A{JB}
6	Pattern A with JPA, followed by Pattern B without JPA	{JA}B
6	Pattern B without JPA, followed by Pattern A with JPA	B{JA}
6	Pattern B with JPA, followed by Pattern A without JPA	{JB}A

As explained earlier in Section 3.2, the assembler intentionally inserted faults in 14 specific locations (seven in each pattern) and there were four different types of faults. Table 4 summarizes the specific faults and associated fault types; raw data results appear in Appendix E.

Table 4: Fault Descriptions/Types

Fault	Fault Description	Fault Type
Fault #1:	Inject kitty 1x2 white piece on far right, facing backward (p.5)	1
Fault #2:	Switch order of 1x1 and 1x3 yellow pieces on left side (p.8)	3
Fault #3:	Switch order of 2x2 and 2x4 yellow pieces on right side (p.9)	3
Fault #4:	Switch order of 1x1 and 1x2 orange pieces on right side (p.14)	3
Fault #5:	Install window backwards (p.15)	4
Fault #6:	Inject kitty 1x2 white piece, facing forward (P.19)	1
Fault #7:	Inject 1x1 eye piece, facing right (P.20)	1
Fault #8:	Make 4th 2x4 black block two 1x4s (p.2)	2
Fault #9:	Don't install red window ahead of white door (p.6)	4
Fault #10:	Inject 1x2 yellow kitty on right side, facing forward (p.10)	1
Fault #11:	Switch order of 2x2 and 2x4 white pieces on right side (p.12)	3
Fault #12:	Inject kitty 1x2 white piece on right side, facing left -- towards QA (p.14)	1
Fault #13:	Make two 1x4 white blocks one 2x4 block (p.14)	2
Fault #14:	Inject 2x2 yellow eye block, face forward (p.16)	1
		1 = Markings
		2 = Incorrect piece(s)
		3 = Wrong order
		4 = Wrong orientation

Figure 2 shows the specific page from the assembly instructions during which Fault #1 was inserted; fault #1 was also fault type 1. The arrow points to the white 1 X 2 Lego™ piece that was chosen for replacement with one of identical shape and color, except with markings on one side.

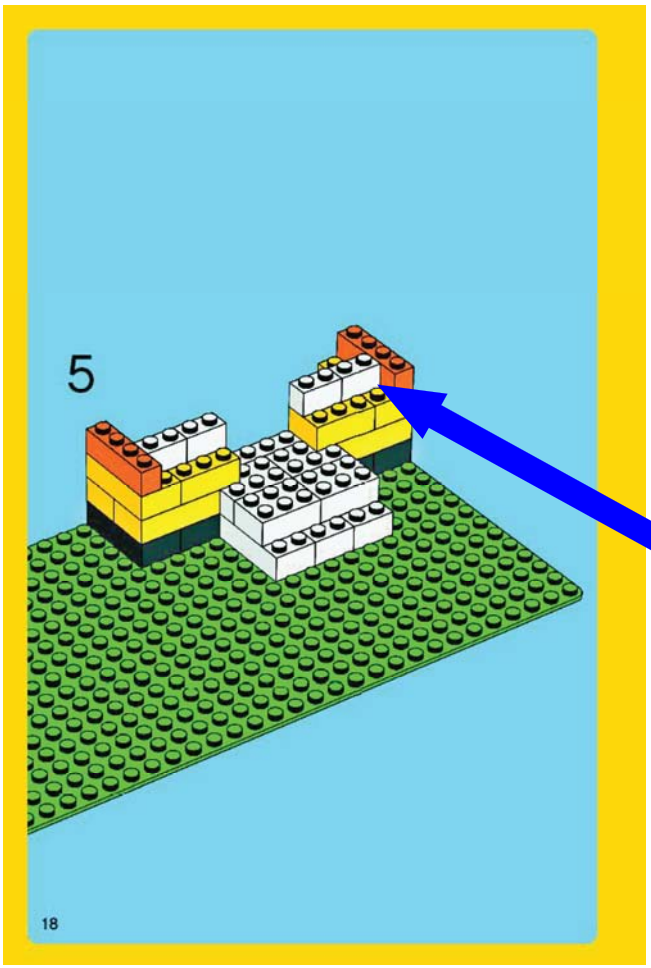


Figure 2: Sample Instructions Page, Depicting Fault #1 (Fault Type 1)

The actual replacement of this piece in an experimental trial is shown by the arrow in Figure 3, where markings are shown on the back side of the part that is facing away from the QA checker. During experimental design it was thought that pieces with markings would be too easily noticeable if they were facing the checker, so it was determined that most of them would face in a different direction. As a counterbalance against these faults then becoming too *difficult* to notice, all pieces with markings were placed face-up on the table during the trial such that the QA checker would have the *opportunity* to notice them.

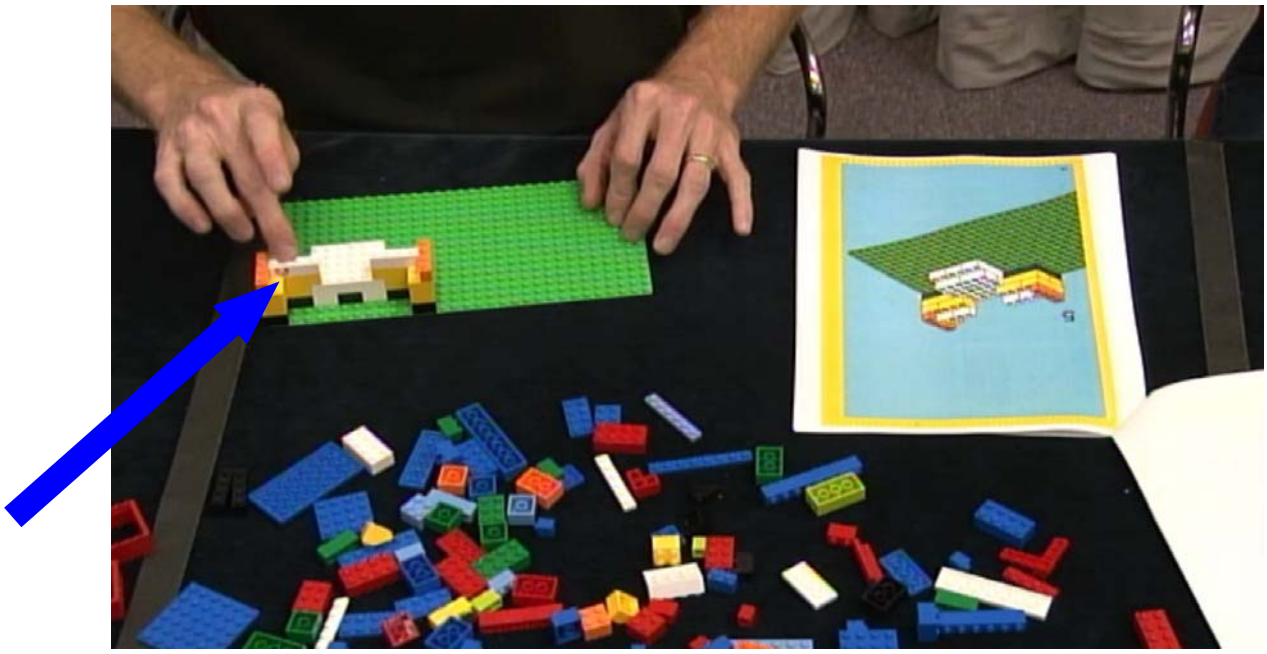


Figure 3: Fault #1 Occurring in an Experimental Trial

All four fault types are illustrated in Appendix F. The complete set of assembly instructions appear in Appendix H (Pattern A) and Appendix I (Pattern B). Note that some pages appear lighter in contrast than others; these pages were purposely altered in order to more accurately identify the black pieces. While most pages depict black pieces it was important to clearly identify those pieces on the page of instructions in which they were installed, to eliminate any confusion. This was deemed an important activity in order to remove the confound of the QA checker's visual acuity, and it was especially important for Fault #8 where a 2 X 4 black piece was replaced by two 1 X 4 black pieces. (This fault, an example of fault type 2, is depicted in Figure F3 and Figure F4.)

3.5 Job Performance Aid

This experiment was designed to focus not on the Lego™ assembly task itself but rather the effectiveness of a JPA for the experimental participant that is performing the role of the QA checker. Prior JPA research (Miller, 1956; Gebhard, 1970; Elliott & Joyce, 1971; Foley, 1972; Clark, 1982; Shriver et al., 1982, Barshi & Healy, 1993) has studied the use of JPAs with both novice and expert users, but there is general agreement that they are most effective as training aids for novice users. Some common themes in the literature proved useful in the development of the checklist for this study:

- **The focus is on the user.** JPA developers should use task analysis techniques to fully understand the job function and the behaviors used when the task is performed (Gebhard, 1970; Clark, 1982; Shriver et al, 1982; Smillie, 1985).
- **Information must be task oriented.** The JPA should identify exactly what the user needs to do using brief, concise, and explicit instructions to aid short term memory. Instructions should be directive and action-specific using simplified and standard language, and contain only information that is necessary (McKenzie, 1958; Gebhard, 1970; Clark, 1982; Shriver et al., 1982; Smillie, 1985; Hallbert et al., 1992; Gawande, 2010)

The JPA for this experiment (see Appendix D) consists of a short, concise, and simple checklist intended to elicit behaviors that would enhance the detection of faults in this experiment. Each of the six checklist items was considered to be essential information, with the following rationale:

- *Your role as an observer is an essential part of this important task. Complex assemblies require a second set of eyes in order to catch any errors*. It is the author's experience that novice QA checkers do not always recognize the importance of simply being an observer for an important task. McKenzie (1958) noted that inspection takes place in a social context where there may be different 'social norms' of rejection, which is consistent with the author noting that subject matter experts who perform the work may sometimes resent a non-expert "checking their work." In addition, Wiener (1975) reminds us of Herzberg's (1966) theory of industrial motivation and quotes Lucaccini et al. (1968) where subjects receiving more positively-toned instructions detected more signals.

- *Pay attention for the following types of error.* Both Clark (1982) and White et al. (2010) recommend that hazards or high-risk concerns have their own specific checklist items, though to avoid confusion the four fault types were addressed in only three statements.
- *Feel free to ask questions about the task at any time. If necessary, ask the assembler to stop until you are comfortable with proceeding*. This statement was intended to build the QA checker's confidence by signaling that they are vital to the assembly process and could stop it at any time. It also grants permission to talk to the assembler and not be considered a distraction.
- *The assembler should not turn to the next page of the instructions without your approval.* This statement was necessary to establish the event horizon at which a

detection becomes a catch. The specific emphasis on catches was kept from the QA checker as it would have introduced a confound.

- *For each page of the instructions, the order of assembly does not matter.* To avoid confusion that could have led to an increase in false alarms, this statement needed to be explicit.
- *The box contains 512 total parts. Some parts will be used and some will not.* This statement eliminates the confound whereby the QA checker could use the presence or absence of parts on the table as a cue.

Clark (1982), Shriver et al. (1982), and Smillie (1985) are emphatic that the key final step in successful development of a JPA is verification and validation with expert users. The process relies heavily on these reviews to identify and correct procedural ambiguities, omissions, and inaccuracies, and depending on the complexity of the JPA the process may require multiple iterations. For the simple and concise checklist in this experiment, the author conducted a pilot study with four QA experts in the Surety Engineering Group at SNL for additional guidance and a final verification of the checklist format. The study was approved (before the pilot study) by both the SNL and Purdue University Institutional Review Boards under protocols SNL1349 and IRB00001150, respectively.

CHAPTER 4. RESULTS AND ANALYSIS

Binary responses for 14 fault observations were recorded for each participant, and are summarized in Table 5. Participant #7 was disqualified due to misunderstanding the pre-test instructions and was replaced with participant #25. The three Pattern B instances where only 6 trials appeared were from incorrect assembly and the planned fault could not be inserted. Note that participant #1 detected every fault in both patterns.

Table 5: Fault Detection Performance, By Participant

Subject #	Pattern A Trials	Pattern A Detections	Pattern B Trials	Pattern B Detections	Percent Detected
1	7	7	7	7	100%
2	7	5	6	5	77%
3	7	4	7	3	50%
4	7	7	6	3	77%
5	7	6	6	4	77%
6	7	5	7	4	64%
8	7	4	7	5	64%
9	7	6	7	5	79%
10	7	5	7	7	86%
11	7	3	7	4	50%
12	7	3	7	4	50%
13	7	3	7	4	50%
14	7	3	7	6	64%
15	7	4	7	4	57%
16	7	4	7	5	64%
17	7	6	7	3	64%
18	7	3	7	4	50%
19	7	4	7	2	43%
20	7	4	7	3	50%
21	7	4	7	5	64%
22	7	4	7	5	64%
23	7	4	7	5	64%
24	7	3	7	3	43%
25	7	4	7	3	50%

Table 6 summarizes the results for each particular fault. The three instances of incorrect assembly are seen again where two of the trials associated with fault #8 and one trial with fault #11 were deemed to be a "no test". It is noteworthy that fault #2, fault #4, and fault #11 were always detected while fault #1, fault #6, fault #7, fault #10, and fault #14 were frequently missed. The members of the latter group (#1, #6, #7, #10, and #14) are all marking faults (fault type 1); however, fault #12 (also fault type 1) was specifically designed to be noticed and, as expected, frequently detected[2].

Table 6: Performance, By Fault Number

Pattern	Fault Number	Fault Type	Number of Trials	Number of Detects	Percent Detected
A	1	1	24	5	21%
A	2	3	24	24	100%
A	3	3	24	23	96%
A	4	3	24	24	100%
A	5	4	24	17	71%
A	6	1	24	6	25%
A	7	1	24	6	25%
B	8	2	22	15	68%
B	9	4	24	21	88%
B	10	1	24	5	21%
B	11	3	23	23	100%
B	12	1	24	20	83%
B	13	2	24	17	71%
B	14	1	24	2	8%

Using the performance summaries in Table 5 and Table 6 as a guide, binary logistic regression (Agresti, 2013) was used to model the probability of detecting a fault as a

[2] This intentionally-noticeable fault was designed to address the vigilance decrement noted by Tsao, Drury,and Morawski (1984) where a subject's expectations of finding another defect appeared to decrease immediately after a defect had been found. However, hypotheses related to vigilance theory were later dropped from this study.

function of the experimental factors. In logistic regression, the probability of detection, π, is assumed to relate to the experimental factors $(x_1, x_2, \cdots, x_p)$ via the general expression $log\left(\frac{\pi(x_1,x_2,\cdots,x_p)}{1-\pi(x_1,x_2,\cdots,x_p)}\right) = f(x_1, x_2, \cdots, x_p)$, where *p* is the number of factors. Note that $log\left(\frac{\pi(x_1,x_2,\cdots,x_p)}{1-\pi(x_1,x_2,\cdots,x_p)}\right)$ is commonly referred to as the *logit* or log odds. Here, we further assume that *f* is a linear combination of the factors. Collectively, the set of values for the experimental factors describes the experimental condition.

Due to the nature of the experiment and the apparent differences in detection ability across faults, separate probability models (see Appendix G) were developed for each pattern. For Pattern A, faults #2 and #4 were excluded from the models because they were always detected and therefore there is no ability to discern an effect of sequence in those cases; for Pattern B, fault #11 was excluded for the same reason. The model form used in each case is:

$$log\left(\frac{\pi(Err(i),\ Seq(j))}{1-\pi(Err(i),\ Seq(j))}\right) = \alpha_0 + \beta_i + \gamma_j, \quad (1)$$

where α_0 represents the log odds at a standard experimental condition (defined by a standard error number and a standard sequence) , β_i reflects the change in log odds when changing the experimental condition from the standard fault number to fault #*i*, anγ_jd reflects the change in log odds when changing the experimental condition from the standard sequence to sequence *j*. In the case of both patterns, the standard sequence is

denoted by pattern A, followed by the JPA instructions and then pattern B, or in study notation, A{JB} (see Table 3). In the case of pattern A, fault #1 is the standard fault number and in the case of pattern B, fault #8 is the standard fault number. The form of the model allows for an assessment of the effect of the JPA given differences in detection ability across fault numbers. The effect of fault number is compensated for by the β_i terms, while the effect of the JPA is associated with the γ_j terms (more on this in Chapter 5).

Minitab (version 16.1.1) was used to estimate the model parameters for each pattern; the complete set of Minitab results appears in Appendix G. Table 7 and Table 8 display the parameter estimates (and associated *p-values*) related with patterns A and B, respectively. The *p-values* measure whether or not the model terms are significantly ($p < 0.05$) different than zero.

Table 7: Logistic Regression Table, Pattern A

Parameter	Estimate	SE Estimate	Z-ratio	P-value
α_0	-2.845	0.810	-3.51	0.000
$\gamma_{B\{JA\}}$	1.792	0.776	2.31	0.021
$\gamma_{\{JA\}B}$	1.999	0.778	2.57	0.010
$\gamma_{\{JB\}A}$	1.578	0.775	2.04	0.042
β_3	4.967	1.218	4.08	0.000
β_5	2.494	0.731	3.41	0.001
β_6	0.251	0.710	0.35	0.724
β_7	0.251	0.710	0.35	0.724

Table 8: Logistic Regression Table, Pattern B

Parameter	Estimate	SE Estimate	Z-ratio	P-value
α_0	0.893	0.593	1.50	0.132
$\gamma_{B\{JA\}}$	0.350	0.637	0.55	0.582
$\gamma_{\{JA\}B}$	-0.615	0.610	-1.01	0.313
$\gamma_{\{JB\}A}$	-0.187	0.612	-0.31	0.760
β_9	1.211	0.776	1.56	0.119
β_{10}	-2.150	0.691	-3.11	0.002
β_{12}	0.870	0.722	1.20	0.228
β_{13}	0.132	0.650	0.20	0.839
β_{14}	-3.227	0.879	-3.67	0.000

For the various experimental conditions, the fitted model can be used to estimate the probability of detection by inverting equation (1) and replacing the parameters with their estimates. For example, suppose that we are interested in the probability of detecting fault #1 (pattern A) assuming the standard sequence A{JB}. Then,

$\hat{\pi}\big(Err(1),\ Seq(\text{A\{JB\}})\big) = \frac{exp(\hat{\alpha}_0)}{1+exp(\hat{\alpha}_0)} = 0.0549$, where $\hat{\alpha}_0 = -2.845$ (from Table 7).

Further suppose that we are interested in the probability of detecting fault #6 (pattern A) assuming the sequence B{JA}. Then, $\hat{\pi}\big(Err(6),\ Seq(\text{B\{JA\}})\big) = \frac{exp(\hat{\alpha}_0+\hat{\beta}_6+\hat{\gamma}_{B\{JA\}})}{1+exp(\hat{\alpha}_0+\hat{\beta}_6+\hat{\gamma}_{B\{JA\}})} = 0.3095$, where $\hat{\alpha}_0, \hat{\beta}_6, and\ \hat{\gamma}_{B\{JA\}}$ are from Table 7. Table 9 and Table 10 list the estimated probabilities of detection (for all experimental conditions) and compare with the observed fraction of faults detected for patterns A and B, respectively. Note that the estimated probability of detection matches reasonably well with the observed fraction detected for all cases. The implication is that the predictors contained in the equation (1) model contain all of the necessary information, meaning that this is an accurate model for the experimental conditions. The same conclusion was reached from Minitab (see Appendix G) using the Pearson (p=0.171), Deviance (p=0.194), and especially Hosmer-

Lemeshow (p=0.725) formal tests where p-values *greater* than 0.05 indicate that there is no evidence for lack-of-fit and that the model is accurate. These results would not be statistically different from increasing the pool of participants except in the case of a *much* larger number that would be required to detect small differences.

Table 9: Estimated Probability of Detection, Pattern A

Fault #	Sequence	Estimated Probability of Detection	Observed Fraction Detected
1	A {JB}	0.055	0.000
1	B {JA}	0.259	0.500
1	{JA} B	0.300	0.167
1	{JB} A	0.220	0.167
3	A {JB}	0.893	1.000
3	B {JA}	0.980	1.000
3	{JA} B	0.984	1.000
3	{JB} A	0.976	0.833
5	A {JB}	0.413	0.500
5	B {JA}	0.809	0.500
5	{JA} B	0.839	0.833
5	{JB} A	0.773	1.000
6	A {JB}	0.069	0.000
6	B {JA}	0.310	0.333
6	{JA} B	0.355	0.500
6	{JB} A	0.266	0.167
7	A {JB}	0.069	0.000
7	B {JA}	0.310	0.333
7	{JA} B	0.355	0.333
7	{JB} A	0.266	0.333

Table 10: Estimated Probability of Detection, Pattern B

Fault #	Sequence	Estimated Probability of Detection	Observed Fraction Detected
8	A {JB}	0.709	0.833
8	B {JA}	0.776	0.800
8	{JA} B	0.569	0.400
8	{JB} A	0.670	0.667
9	A {JB}	0.891	0.833
9	B {JA}	0.921	0.833
9	{JA} B	0.816	0.833
9	{JB} A	0.872	1.000
10	A {JB}	0.221	0.167
10	B {JA}	0.288	0.333
10	{JA} B	0.133	0.333
10	{JB} A	0.191	0.000
12	A {JB}	0.854	1.000
12	B {JA}	0.892	0.667
12	{JA} B	0.759	0.833
12	{JB} A	0.829	0.833
13	A {JB}	0.736	0.667
13	B {JA}	0.798	0.833
13	{JA} B	0.601	0.500
13	{JB} A	0.698	0.833
14	A {JB}	0.088	0.000
14	B {JA}	0.121	0.333
14	{JA} B	0.050	0.000

CHAPTER 5. DISCUSSION

The effect of the JPA can be deduced by considering the complete set of estimated γ_j terms in the model, which represent the 4 test conditions shown in Table 3. In the case of pattern A (see Table 7), note that $\hat{\gamma}_{B\{JA\}}$, $\hat{\gamma}_{\{JA\}B}$, and $\hat{\gamma}_{\{JB\}A}$ are all statistically significantly non-zero (each with a p-value less than 0.05) and positive, thus implying increased probability of detection versus the standard sequence A{JB}. This means that faults in Pattern A were detected *less frequently* in the standard A{JB} sequence, when the Pattern A assembly was observed first and the JPA appeared in the second trial on Pattern B. This suggests that the JPA may have had only a limited effect in this experiment. Further, the values of the three γ parameter estimates in Table 7 are statistically indistinguishable from one another (the standard error estimates are essentially identical). Thus, the probability of detection of Pattern A faults is impacted by *both* the sequence of observation (whether or not the assembly of pattern A was observed before the assembly of pattern B) *and* whether or not a JPA was present.

The only distinguishable effect of the JPA is its presence in a *3-way interaction* between sequence, presence/absence of a JPA, and Pattern A. If there was an effect from *only* the JPA, the $\hat{\gamma}_{B\{JA\}}$ and $\hat{\gamma}_{\{JA\}B}$, terms in Table 7 would be statistically significant and positive while $\hat{\gamma}_{\{JB\}A}$ would be near zero. These results for Pattern A further suggest that the

learning effect has a greater impact than the JPA. In contrast, the effects of sequence and/or presence/absence of a JPA on the probability of detection for pattern B (see Table 8) were not observed (i.e., $\hat{\gamma}_{B\{JA\}}$, $\hat{\gamma}_{\{JA\}B}$, and $\hat{\gamma}_{\{JB\}A}$ were all statistically indistinguishable from zero with non-significant p-values).

As seen in Table 6, marking faults (fault type 1) were frequently missed and therefore dominate the 3-way interaction term in the results for Pattern A. This can be seen in the β_3 term in Table 7 with the statistically significant p-value < 0.0005 and a large positive value for the estimate (4.967). This indicates that for Pattern A, participants were *much* more likely to detect fault #3 (and fault type 3, which is an incorrect-order fault) compared to fault #1 (the default condition, as well as a marking fault). Similarly, the β_{10} term for Pattern B in Table 8 is statistically significant (p=0.002) with a large *negative* estimate (-2.150) and signifies that the marking fault #10 was much *less* likely to be detected than the incorrect-piece fault (fault type 2) in the default condition. This suggests that the reason Pattern A appears in the 3-way interaction term is because it has more marking errors (even when considering fault #12, which was designed to be noticed) and thus gives a better opportunity to detect differences in the probability of detection between the different fault types.

Given these results, the primary null hypothesis was not fully rejected:

H_1: The presence of a JPA has no effect on the detection of faults in the quality assurance role.

The effect of the JPA was not statistically significant on its own but rather as part of a 3-way interaction between sequence of observation, presence of a JPA, and Pattern A. Thus, the additional null hypotheses were only partially accepted:

H_2: There is no difference in the detection of faults with a JPA than without a JPA.
H_3: The order of presentation of the JPA has no effect on the detection of faults in the quality assurance role.
H_4: The order of presentation of the different patterns has no effect on the detection of faults in the quality assurance role.

What may be the most intriguing discovery in this study, however, is the ability to *reject* the final null hypothesis:

H_5: The probability of detection for each fault is equal.

In hindsight, one weakness in the experimental design is that faults were not evenly distributed throughout the individual trial for each pattern, either by elapsed time or number of pieces of assembly. This might have been negligible if the assumption of average probability of detection across all faults was realistic; however, this assumption was shown to be demonstrably *wrong* and therefore acted as a confound influencing the expected results. Clearly, some fault types in this study were more difficult to detect than other fault types. The discovery that there are different probabilities of detection for

different types of error, using this study's experimental design and within the verification context, may be the most useful result from this experiment.

Returning to the original motivation for the study provides a reminder that this research project is not just an experiment about Lego™ pieces and JPAs, but also a study of QA practices. In order for QA practitioners to effectively use tools to control risk they must be both well designed and well understood by the designer and the user. Stammers et al. (1990) suggests that task analysis methodologies themselves are limited by the difficulty of accurately predicting behavior due to the influence of cognitive factors that cannot be easily observed or modeled, thus even proper JPA development using task analyses face the same limitations. If this study were designed to detect only the main effect of a JPA on performance, then the results would have been disappointing and perhaps no interaction with other independent variables would have been discovered. These findings illustrate that this specific JPA format (i.e., a checklist) is not necessarily an effective control to improve the performance of concurrent dual verification activities. Past findings (Clark, 1982; Shriver et al., 1982; Smillie, 1985) echo the conclusion of this research that no single JPA format is best for all circumstances. It should also be noted that the verification techniques appearing in many guidebooks, such as the DOE *Human Performance Handbook* (DOE, 2009) mentioned earlier in Chapter 2, are presented as being effective without any qualifications as to whether they are actually useful or conditions under which they must be applied in order to be beneficial. This study found, however, that concurrent dual verification may not always be as effective a technique as expected or desired, either with or without a JPA.

CHAPTER 6. CONCLUSION AND FUTURE RESEARCH

The research presented in this thesis is the first known example of evidence-based job performance aid (JPA) use in a quality assurance setting. This study used an assembly task and checklist format as the vehicle for the experiment, and incorporated previously recognized design principles from the literature into the configuration of the JPA. The design of the task included some simplifying assumptions regarding an average probability of detection between different types of error, and this led to the primary research hypotheses unable to be proved. Nevertheless, there were 3 important findings.

First, this study created a testing methodology sensitive enough to detect differences in the effects on performance between the 1) sequence of observation of patterns, 2) presence/absence of a JPA, and 3) Pattern A. The author reflects that if the *main effect* of a JPA on performance of a concurrent dual verification task were easily identifiable then it would likely have been detected long ago, though it is worth noting that there has not been any research located on concurrent dual verification either with or without the use of JPAs. Second, the results indicate that concurrent dual verification *itself* is not necessarily as effective a control as presented in the literature (however, some error detection is probably better than none in most cases). One cannot assume that having a QA checker in place will have the desired impact on error detection or mitigation,

especially for specific types of errors. Third, the assumption of average probability of detection between different types of error may have been demonstrably wrong but the error detection probability could be empirically *verified* within the error construct created during this experiment. Because the probability of detection of different error types has not previously been studied in this context, it is hoped that these results are well-received by a wide community of practitioners. In addition, the subtle and complex interactions of JPA design, error type, and base error probability identified in this study help to clarify the lack of clear findings in past research.

These results can be used as a cautionary tale for processes and tools that QA practitioners use to prevent or mitigate human errors. Once an error is found, the simple conclusion is usually to blame the assembler (or checker) without understanding the complexity of tasks that both are being asked to perform. Caldwell (2008) reminds us to look for patterns of events from multiple causal factors, but this study contributes another dimension by recognizing that different types of error may require different mitigation techniques. The common risk matrix (Cox, 2008) describes the likelihood of errors occurring against their consequences, but this is *independent* of the probability of those errors being detected during a concurrent checking task. The likelihood of an error occurring can never be zero but if the probability of an error being *detected* can be improved, then an effective use of QA controls might be to improve the chances that catastrophic errors do not go unnoticed. The recognition in this study that different types of error have different probabilities of detection, which could then be inserted into a

checklist as behavioral cues, might be highly useful as a leverage to the success of both JPA design and concurrent dual verification activities.

Both the methodology and results of this study are an effective baseline from which to launch future research activities. Results suggest that a learning effect overshadowed the presence of the JPA, therefore a redesigned study to counterbalance this effect might be useful. Repeating the same experiment with a uniform fault type, which would likely provide a more constant probability of detection for all faults, may also yield intriguing results. A focus on marking faults, which can be considered signals, might provide an attractive opportunity. Since incorrect markings may fall below the signal detection threshold (Swets, 1964; Baker, 1975), such a focused study might extend this essential paradigm of vigilance theory (e.g., Mackie, 1964; Kibler, 1965; Stanislaw, 1990; Caldwell, 1999). If a similar experiment equally spaced the faults (possibly by elapsed time) throughout the assembly task, then perhaps a vigilance decrement could be studied in a number of ways. It was mentioned previously that fault #12 was designed to address findings by Tsao, Drury, and Morawski (1979) and a later study by Tsao and Wang (1984) examined faults of different difficulty, suggesting that these faults may have different probabilities of detection. These and other vigilance studies may benefit from the recognition of different probabilities of detection for different types of error, both within and outside of the concurrent dual verification construct. Altmann (2002) has fundamental areas of overlap with the current research by examining a performance decline within the current task as memory decays (or loses activation) to make room for encoding the next task. Similarly, Altmann and Trafton (2007) discuss the effects of

delayed response and performance recovery after interruptions. While this study was unable to directly test the activation or interruption recovery concept, given that the presence of a JPA is not a repetitive intervention, a future experiment with multiple JPAs could address this activation model. Finally, it is worth noting that the QA checkers in this experiment used a simple checklist as the JPA, and future studies may have different results with other JPA formats. Perhaps as a result of such studies, optimal JPA formats for different fault types may emerge.

REFERENCES

REFERENCES

Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). New York: John Wiley.

Allen Jr, J.P., & Rankin, W.L. (1997). Use of the Maintenance Error Decision Aid (MEDA) to Enhance Safety and Reliability and Reduce Costs in the Commercial Aviation industry. In *The 10th Annual FAA Conference on Aviation Maintenance and Inspection.*

Altmann, E.M. (2002). Functional Decay of Memory for Tasks. *Psychological Research, 66*(4), 287-297.

Altmann, E.M., & Trafton, J.G. (2007). Timecourse of recovery from task interruption: Data and a model. *Psychonomic Bulletin & Review*, *14*(6), 1079-1084.

Baddeley, A.D., & Colquhoun, W.P. (1969). Signal Probability and Vigilance: A Reappraisal of the 'Signal-Rate' Effect. *British Journal of Psychology, 60*(2), 169-178.

Baker, E.M. (1975). Signal Detection Theory Analysis of Quality Control Inspector Performance. *Journal of Quality Technology, 7*, 62-71.

Barshi, I., & Healy, A.F. (1993). Checklist Procedures and the Cost of Automaticity. *Memory & Cognition, 21*(4), 496-505.

Booher, H.R. (1975). Relative Comprehensibility of Pictorial Information and Printed Words in Proceduralized Instructions. *Human Factors, 17*(3), 266-277.

Booher, H.R. (1978). *Job Performance Aids: Research and Technology State-of-the-Art*. Technical Report NPRDC-TR-78-26. San Diego, CA: Navy Personnel Research and Development Center.

Caldwell, B.S. (1999). Dynamics of Supervisory Control Task Performance: SCAMPI Project Summary. In *Automation Technology and Human Performance: Current Research and Trends; Proceedings of the Third Conference on Automation Technology and Human Performance*, Norfolk, VA, March 25-28, 1998, 213-217.

Caldwell, B.S. (2008). Tools for Developing a Quality Management Program: Human Factors and Systems Engineering Tools. *International Journal of Radiation Oncology * Biology* Physics, 71*(1), S191–S194.

Chapanis, A. (1970). Systems Staffing. In K.B. DeGreene (Ed.). *Systems Psychology*, 357-382. New York: McGraw-Hill.

Cheng, P.G., Hecht, T.C., & Bohner, J.J. (2009) *Why Satellites Fail: Lessons for Mission Success*. Report TOR-2009(8617)-8704. El Segundo, CA: The Aerospace Corporation.

Clark, J.E. (1982). *JPAs: An Actual Experience*. No. UNI-SA-96; CONF-820223-1. Richland, WA: UNC Nuclear Industries, Inc.

Cox, L.A., Jr. (2008). What's Wrong With Risk Matrices? *Risk Analysis, 28*(2), 497-512.

Davies, I.K. (1973). Task analysis. *AV Communication Review, 21*(1), 73-86.

Dhillon, B.S. (2008). *Reliability Technology, Human Error, and Quality in Health Care*. CRC Press.

DOE. (1993). *Guide to Good Practices for Independent Verification.* DOE-STD-1036-93. Washington, DC: United States Department of Energy Technical Standards Program.

DOE. (2009). *Human Performance Improvement Handbook, Volume 2: Human Performance Tools for Individuals, Work Teams and Management.* DOE-HDBK-1028-2009. Washington, DC: United States Department of Energy Technical Standards Program.

Drury, C.G. (1982). Improving Inspection Performance. In G. Salvendy (Ed.), *Handbook of Industrial Engineering*. New York: John Wiley.

Elliott, E. (1960). Perception and Alertness. *Ergonomics, 3*(4), 357-364.

Elliott, T.K., and Joyce, R.P. (1971) An Experimental Evaluation of a Method for Simplifying Electronic Maintenance. *Human Factors, 13*(3), 217-227.

Foley, J.P., Jr. (1972). *Description and Results of the Air Force Research and Development Program for the Improvement of Maintenance Efficiency*. Technical Report AFHRL-TR-72-72. Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory.

Folley, J.D., Jr. & Shettel, H.H. (1962) *Tryout of a Preliminary Procedure for Systematically Designing Performance Aids.* Technical Report MRL-TDR62-20. Wright-Patterson AFB, OH: Behavioral Sciences Laboratory.

Folley, J.D., Jr. (1961). *A Preliminary Procedure for Systematically Designing Job Performance Aids*. Technical Report 61-550. Wright-Patterson AFB, OH: Aeronautics Systems Division, Air Force Systems Command.

Folley, J.D., Jr., Joyce, R.P., Mallory, W.L., & Thomas, D.L. (1971). *Fully Proceduralized Job Performance Aids: Volume 1. Draft Specification for Organizational*

Maintenance. Technical Report AFHRL-TR-71-53. Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory.

Fox, J.G. (1973). Recent Human Factors Contribution to Enhancing Industrial Quality Control. *Behaviorometric, 3*(2), 99-118.

Gagne, R.M. (1962). Human Functions in Systems. In R.M. Gagne (Ed.) *Psychological Principles in System Development*, 35-73. New York: Holt, Rinehart, and Winston.

Gawande, A. (2010). *The Checklist Manifesto: How to Get Things Right*. New York: Picador.

Gebhard, R.M. (1970). *Development of a Training Program and Job Aids for Maintenance of Electronic Communication Equipment.* Technical Report HUMRRO-TR-70-19. Alexandria, VA: Human Resources Research Organization.

Goff, J., Schlesinger, R., & Parlog, J. (1969). *Project PIMO Final Report. Volume 2. PIMO Test Summary*. Chatsworth, CA: Serendipity, Inc.

Guttman, L. (1944). A Basis for Scaling Qualitative Data. *American Sociological Review, 9*(2), 139-150.

Hallbert, B.P., Rodriguez, M.A., and Harbour, G.L. (1992). *Development of Job Performance Aids to Increase Human Performance Reliability: A Case Study in the Evaluation of Human Factors Principles.* Report RFP-4496; CONF-921048-1. Golden, CO: EG&G Rocky Flats, Inc.

Harris, Douglas H. (1969). The Nature of Industrial Inspection. *Human Factors, 11*(2), 139-148.

Haynes, A.B., Weiser, T.G., Berry, W.R., Lipsitz, S.R., Breizat, A.H.S., Dellinger, E.P., Herbosa, T., Joseph, S., Kibatala, P.L., Lapitan, M.C.M., Merry, A.F., Moorthy, K., Reznick, R.K., Taylor, B., & Gawande, A.A. (2009). A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *New England Journal of Medicine*, *360*(5), 491-499.

Helmreich, R.L. (2000). On Error Management: Lessons From Aviation. *British Medical Journal, 320*(7237), 781-785.

Helmreich, R.L., & Merritt, A.C. (2001). *Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences*. Farnham, UK: Ashgate Publishing.

Helmreich, R.L., & Schaefer, H.G. (1994). Team Performance in the Operating Room. In M.S. Bogner (Ed.) *Human Error in Medicine.* Hillside, NJ: Lawrence Erlbaum.

Herzberg, F.I. (1966). *Work and the Nature of Man*. New York: World Publishing.

Joyce, R.P., Chenzoff, A.P., Mulligan, J.F., & Mallory, W.J. (1973a). *Fully Proceduralized Job Performance Aids: Vol. I. Draft Military Specifications for Organizational Level and Intermediate Maintenance*. Technical Report AFHRL-TR-73–43(I). Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory.

Joyce, R.P., Chenzoff, A.P., Mulligan, J.F., & Mallory, W.J. (1973b). *Fully Proceduralized Job Performance Aids: Vol. II. Handbook for JPA Developers*. Technical Report AFHRL-TR-73–43(II). Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory.

Kemeny, J.G. (1979). *The Need for Change, the Legacy of TMI: Report of the President's Commission on the Accident at Three Mile Island*. Washington, DC: U.S. Government Printing Office.

Kibler, A.W. (1965). The Relevance of Vigilance Research to Aerospace Monitoring Tasks. *Human Factors, 7*(2), 93-99.

Lucaccini, L F., Freedy, A., & Lyman, J. (1968). Motivational Factors in Vigilance: Effects of Instructions on Performance in a Complex Vigilance Task. *Perceptual and Motor Skills, 26*(3), 783-786.

Mackie, R.R. (1964). *Six Years of Research on Human Factor Problems in ASW: A Summary.* Technical Report 206-28. Los Angeles, CA: Human Factors Research Inc.

Mackworth, N.H (1950). Researches on the Measurement of Human Performance. *Medical Research Council Special Report Series,* No. 268.

Mager, R.F. (1962). *Preparing Objectives for Programmed Instruction*. Belmont, CA: Fearon Publishers, Inc.

Mattson, R.J. (1989). *An Assessment of Criticality Safety at the Department of Energy, Rocky Flats Plant*. SCIEDOE-201-89.

McGrath, J.J., & Harabedian, A. (1961). Signal Detection as Function of Intersignal Interval Duration. In *Human Factors Problems in Antisubmarine Warfare* (Technical Report #9). Los Angeles, CA: Human Factors Research, Inc.

McKenzie, R.M. (1958). On the Accuracy of Inspectors. *Ergonomics 1*(3), 258-272.

Meilinger, P.S. (2004). When the Fortress Went Down. *Air Force Magazine, 87*, 78-82.

Miller, R.B. (1953). *A Method for Man-Machine Task Analysis*. Technical Report 53-137. Wright-Patterson AFB, OH: Wright Air Development Center.

Miller, R.B. (1956). *A Suggested Guide to the Preparation of Handbooks of Job Instructions*. Pittsburgh, PA: American Institute for Research.

Murray, S.A., & Caldwell, B.S. (1996). Human Performance and Control of Multiple Systems. *Human Factors, 38*(2), 323-329.

Newman, S.E. (1957). *On Identifying Effective Handbook Techniques* (ML-TM-57-26). Lowry Air Force Base, CO: Maintenance Laboratory, Air Force Personnel and Training Research Center.

Norman, D.A. (1981). Categorization of Action Slips. *Psychological Review, 88*(1), 115.

Norman, D.A. (1983). Design Rules Based on Analyses of Human Error. *Communications of the ACM, 26*(4), 254-258.

Oberhettinger, D. (2012). Assuring that Lessons Learned Critical to Mission Success Get Used. In *2012 IEEE Aerospace Conference*, March 4-9, Paper #1177. Institute of Electrical and Electronics Engineers.

Official: Genesis Pre-Launch Test Skipped. (2006, January 7) *Space.com* Retrieved from http://web.archive.org/web/20060110200423/http://www.space.com/missionlaunches/ap_060107_genesis_update.html

Posner, M.I. (1964). Information Reduction in the Analysis of Sequential Tasks. *Psychological Review, 71*(6), 491-504.

Rasch, G. (1960). *Studies in Mathematical Psychology: Probabilistic Models for Some Intelligence and Attainment Tests*. Oxford, England: Nielsen & Lydiche.

Rasmussen, J. (1982). Human Errors: A Taxonomy for Describing Human Malfunction in Industrial Installations. *Journal of Occupational Accidents, 4*(2), 311-333.

Rasmussen, J., & Vicente, K.J. (1989). Coping with Human Errors Through System Design: Implications for Ecological Interface Design. *International Journal of Man-Machine Studies, 31*(5), 517-534.

Reason, J. (1990). *Human Error*. London, UK: Cambridge University Press.

Russians Miswire Satellite. (1996, March 21) *The Moscow Times*. Retrieved from http://www.themoscowtimes.com/news/article/russians-miswire-satellite/326829.html

See, J.E. (2012) *Visual Inspection: A Review of the Literature*. SAND2012-8590. Albuquerque, NM: Sandia National Laboratories.

Shriver, E L., Fink, C.D., & Trexler, R.C. (1959). *Increasing Electronics Maintenance Proficiency Through Cue-Response Analysis*. Alexandria, VA: Human Resources Research Organization.

Shriver, E.L. (1960). *Determining Training Requirements for Electronic System Maintenance: Development and Test of a New Method of Skill and Knowledge Analysis*. Technical Report 63. Alexandria, VA: Human Resources Research Organization.

Shriver, E.L. (1975). *Fully Proceduralized Job Performance Aids: Guidance for Performing Behavioral Analyses of Tasks*. Falls Church, VA: URS/Matrix Research Company.

Shriver, E.L., Zach, S.E., and Foley Jr, J.P. (1982). *Test of Job Performance Aids for Power Plants*. No. EPRI-NP-2676. Alexandria, VA: Kinton, Inc.

Smillie, R.J. (1978). Continuing Research in Job Performance Aids: The Interaction of Speed Stress and Media. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 22*(1), 502-506. SAGE Publications.

Smillie, R.J. (1985). Design Strategies for Job Performance Aids. In T.M. Duffy & R.W. Waller (Eds.) *Designing Usable Texts*, 213-241. Academic Press, Inc.

Smillie, R.J., & Ayoub, M.A. (1980). Job Performance Aids: Evaluation of Design Alternatives Via Network Simulation. *Ergonomics, 23*(4), 319-339.

Stammers, R.B., Carey, M.S., & Astley, J.A. (1990). Task Analysis. In: J.R. Wilson & E.N. Corlen (Eds.) *Evaluation of the Human Work. A practical Ergonomics Methodology*, 134-160. London, UK: Taylor & Francis Ltd.

Stanislaw, Harold. (1995). Effect of Type of Task and Number of Inspectors on Performance of an Industrial Inspection-type Task. *Human Factors, 37*(1), 182-192.

Stevens-Adams, S., Cole, K., Wenner, C., & Dornburg, C. (2013). *The Effectiveness of Verification in Industrial Applications*. Unpublished manuscript, Sandia National Laboratories.

Swets, J.A. (1964). *Signal Detection and Recognition by Human Observers*, New York: Wiley & Sons.

Thomson, R.M. (1972). Design of Multi-Man Machine Work Areas. In H.P. Van Cott and R.G. Kinkade (Eds.) *Human Engineering Guide To Equipment Design* (Chapter 10). Washington, DC: U.S. Government Printing Office.

Tsao, Y.C., & Wang, T.G. (1984). Inspectors' Stopping Policy After Fault Detection. *Human Factors, 26*(6), 649-657.

Tsao, Y.C., Drury, C.G., & Morawski, T.B. (1979). Human Performance in Sampling Inspection. *Human Factors, 21*(1), 99-105.

White, R.E., Trbovich, P.L., Easty, A.C., Savage, P., Trip, K., & Hyland, S. (2010). Checking it Twice. *Quality and Safety in Health Care, 19*(6), 562-567.

Wiener, E.L. (1975). Individual and Group Differences in Inspection. In C.G. Drury & J.G. Fox (Eds.) *Human Reliability in Quality Control*, 101-122. Halsted Press.

APPENDICES

Appendix A Pattern A

Figure A1: Pieces from Pattern A (Lego™ Kit 5508)

Part ID	# of Parts	Part ID	# of Parts	Part ID	# of Parts
4528139	1	4521943	1	611226	2
4528164	3	300901	1	303726	9
4563682	2	4181142	2	4180814	2
303221	1	300101	1	300924	4
371021	3	300401	5	300124	6
300421	2	300501	2	301024	2
300521	2	300201	4	300224	7
4552353	2	4181144	2	362224	2
4153827	2	300126	2	300324	5
4118827	2	300326	3	235724	4
4121739	2	303926	1	300424	4
4173805	2	4121966	1	300524	2
4520842	6			**Total:**	**104**

Figure A2: Lego™ Kit 5508 Contents

Appendix B Pattern B

Figure B1: Pieces from Pattern B (Lego™ Kit 6167)

Part ID	# of Parts	Part ID	# of Parts	Part ID	# of Parts
621628	2	300126	4	4247139	1
4220632	2	301026	2	300401	4
4124107	2	3004626	1	300501	2
4153827	4	302226	1	300724	2
4118827	2	306226	4	300824	2
4211183	2	4113220	1	4181143	2
413221	3	395701	1	300824	2
365921	2	4218637	1	301024	1
4162384	1	300701	2	300124	7
4144003	1	300801	2	300224	12
304321	1	4181142	3	362224	4
4157124	2	301001	2	300324	9
300421	4	300901	6	300424	2
4129539	9	300201	4	300524	2
303921	6	300301	4	407024	1
303721	8	300101	6	3005744	1
4217795	1			**Total:**	**150**

Figure B2: Lego™ Kit 6167 Contents

Appendix C Incorrect Parts

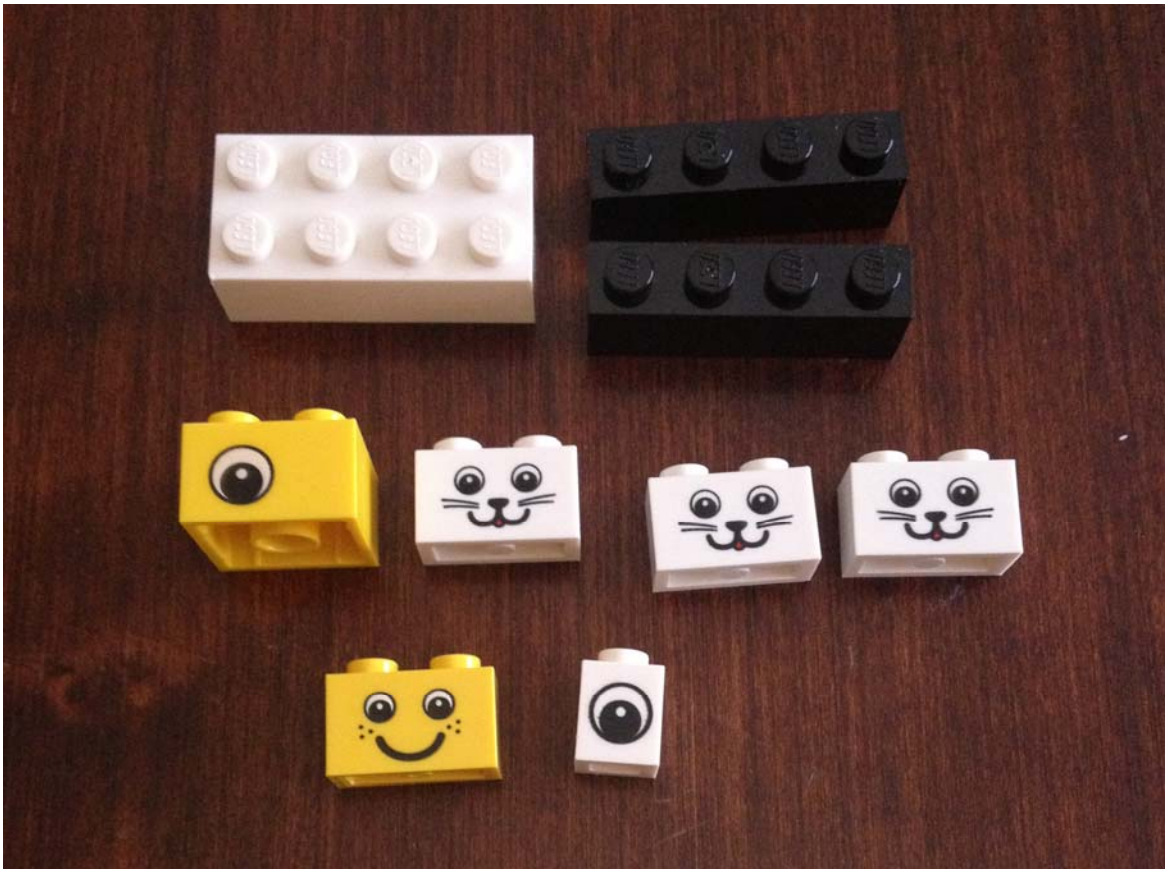


Figure C1: Parts Used for Faults

Appendix D Job Performance Aid

Job Performance Aid

For the Quality Assurance Role

- Your role as an observer is an essential part of this important task. Complex assemblies require a second set of eyes in order to catch any errors.
- Pay attention for the following types of error:
  - An incorrect piece is installed, meaning that it is either the wrong size, wrong color, or wrong markings
  - The correct piece is installed, but in the wrong orientation
  - The correct piece is installed, but in the wrong location
- Feel free to ask questions about the task at any time. If necessary, ask the assembler to stop until you are comfortable with proceeding.
- The assembler should not turn to the next page of the instructions without your approval.
- For each page of the instructions, the order of assembly does not matter.
- The box contains 512 total parts. Some parts will be used and some will not.

Appendix E Raw Data

						First Pattern																															
	Male = 14 Female = 10					Video Start Time		Miss #1 Elapsed Time		Error Type			Miss #2 Elapsed Time		Error Type			Miss #3 Elapsed Time		Error Type			Miss #4 Elapsed Time		Error Type			Video End Time									
Participant	Gender	Age	Presentation Order	Pattern A First	JPA First	Minute	Second	Minute	Second	Markings	Incorrect Order	Orientation	Minute	Second	Markings	Incorrect Order	Orientation	Minute	Second	Markings	Incorrect Order	Orientation	Minute	Second	Markings	Incorrect Order	Orientation	Minute	Second	Total Pattern Time	Miss #1 Actual Time	Miss #2 Actual Time	Miss #3 Actual Time	Miss #4 Actual Time				
1	M	49	B without JPA, A with JPA			2	20																					27	6	24:46								
2	F	40	A with JPA, B without JPA	1	1	2	53	6	35	1			21	24	1													21	33	18:40	03:42	18:31						
3	M	35	A without JPA, B with JPA	1		3	10	6	32	1			19	36	1			20	14	1								20	22	17:12	03:22	16:26	17:04					
4	M	28	A with JPA, B without JPA	1	1	3	18																					20	16	16:58								
5	F	54	B without JPA, A with JPA			2	42	14	28	1			21	24	1													26	8	23:26	11:46	18:42						
6	F	30	B with JPA, A without JPA		1	2	31	11	24	1			16	15	1			18	15	1								22	37	20:06	08:53	13:44	15:44					
7 (DQ)	F	56	A without JPA, B with JPA																																			
8	M	30	B with JPA, A without JPA		1	4	10	14	47	1			23	41	1													30	31	26:21	10:37	19:31						
9	F	35	B with JPA, A without JPA		1	2	50	13	54	1			20	30	1													24	12	21:22	11:04	17:40						
10	M	24	B without JPA, A with JPA			2	5																					22	14	20:09								
11	F	54	A with JPA, B without JPA	1	1	3	18	6	14	1			14	45			1	17	33	1			18	0	1			18	10	14:52	02:56	11:27	14:15	14:42				
12	M	43	A without JPA, B with JPA	1		2	43	5	23	1			13	43			1	15	58	1			16	24	1			16	31	13:48	02:40	11:00	13:15	13:41				
13	M	49	B without JPA, A with JPA			2	22	9	49	1			13	23	1			15	12	1								18	53	16:31	07:27	11:01	12:50					
14	M	25	A without JPA, B with JPA	1		1	57	4	26	1			12	2			1	13	50	1			14	15	1			14	26	12:29	02:29	10:05	11:53	12:18				
15	F	56	B with JPA, A without JPA		1	2	4	2	48		1		10	22	1			16	5	1								19	24	17:20	00:44	08:18	14:01					
16	F	39	B with JPA, A without JPA		1	3	36	13	5	1			19	24	1													23	30	19:54	09:29	15:48						
17	F	37	A with JPA, B without JPA	1	1	3	38	7	54	1																		20	40	17:02	04:16							
18	M	47	A without JPA, B with JPA	1		2	26	5	39	1			12	45			1	14	42	1			15	8	1			15	21	12:55	03:13	10:19	12:16	12:42				
19	M	48	A with JPA, B without JPA	1	1	3	9	6	30	1			16	51	1			17	29	1								17	46	14:37	03:21	13:42	14:20					
20	M	65	B without JPA, A with JPA			2	44	14	8	1			19	24	1			19	30		1		21	15	1			25	44	23:00	11:24	16:40	16:46	18:31				
21	M	46	A without JPA, B with JPA	1		2	20	4	42	1			14	2	1			14	26	1								14	36	12:16	02:22	11:42	12:06					
22	M	55	A with JPA, B without JPA	1	1	3	7	5	42	1			16	2	1			16	25	1								16	33	13:26	02:35	12:55	13:18					
23	F	34	A without JPA, B with JPA	1		2	42	5	49	1			15	53	1			16	27	1								16	36	13:54	03:07	13:11	13:45					
24	M	53	B without JPA, A with JPA			3	50	4	38		1		8	8			1	12	42	1			18	51	1			22	22	18:32	00:48	04:18	08:52	15:01				
25	F	58	B with JPA, A without JPA		1	5	2	6	50	1			14	19	1			19	8		1		21	3	1			24	40	19:38	01:48	09:17	14:06	16:01				

Figure E1: First Pattern Raw Data (1 of 2)

						First Pattern																	
	Male = 14 Female = 10					Misses																	
Participant	Gender	Age	Presentation Order	Pattern A First	JPA First	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	False Alarms	Mistakes	First Half Misses	Second Half Misses	Notes
1	M	49	B without JPA, A with JPA										0	0	0	0	0	0	0	1	1	0	0	Test subject is red/green color blind
2	F	40	A with JPA, B without JPA	1	1	1	0	0	0	0	0	1								0	0	1	1	
3	M	35	A without JPA, B with JPA	1		1	0	0	0	0	1	1								0	1	1	2	
4	M	28	A with JPA, B without JPA	1	1	0	0	0	0	0	0	0								0	0	0	0	
5	F	54	B without JPA, A with JPA											0	1	0	0	0	1	0	1	1	1	Only 6 errors, forgot to perform p.2 fault correctly
6	F	30	B with JPA, A without JPA		1								0	0	1	0	1	0	1	0	3	1	2	
7 (DQ)	F	56	A without JPA, B with JPA																					Disqualified due to misunderstanding pre-trial instructions
8	M	30	B with JPA, A without JPA		1								0	0	1	0	0	0	1	0	1	1	1	
9	F	35	B with JPA, A without JPA		1								0	0	1	0	0	0	1	0	1	0	2	
10	M	24	B without JPA, A with JPA										0	0	0	0	0	0	0	0	0	0	0	
11	F	54	A with JPA, B without JPA	1	1	1	0	0	0	1	1	1								0	0	1	3	
12	M	43	A without JPA, B with JPA	1		1	0	0	0	1	1	1								0	0	1	3	Miss #2 became a catch at 14:26
13	M	49	B without JPA, A with JPA										0	0	1	0	1	0	1	0	0	1	2	
14	M	25	A without JPA, B with JPA	1		1	0	0	0	1	1	1								0	0	1	3	
15	F	56	B with JPA, A without JPA		1								1	0	1	0	0	0	1	2	0	2	1	Norman's mode error, p.2, 2:39 elapsed on video; pointed out a missing piece even before I had a chance to lay it down - and then missed the error that I inserted for that very piece, at 2:48 elapsed
16	F	39	B with JPA, A without JPA		1								0	0	1	0	0	0	1	0	0	1	1	
17	F	37	A with JPA, B without JPA	1	1	1	0	0	0	0	0	0								1	0	1	0	Miss #1 became a catch at 20:09
18	M	47	A without JPA, B with JPA	1		1	0	0	0	1	1	1								0	1	1	3	
19	M	48	A with JPA, B without JPA	1	1	1	0	0	0	0	1	1								0	0	1	2	
20	M	65	B without JPA, A with JPA										0	0	1	0	1	1	1	0	0	1	3	Miss #2 became a catch at 19:42
21	M	46	A without JPA, B with JPA	1		1	0	0	0	0	1	1								0	1	1	2	
22	M	55	A with JPA, B without JPA	1	1	1	0	0	0	0	1	1								0	0	1	2	
23	F	34	A without JPA, B with JPA	1		1	0	0	0	0	1	1								0	0	1	2	
24	M	53	B without JPA, A with JPA										1	1	1	0	0	0	1	0	0	3	1	Miss #2 became a catch at 9:12
25	F	58	B with JPA, A without JPA		1								1	0	1	0	0	1	1	1	0	2	2	
						11	0	0	0	4	9	10	3	1	10	0	3	2	10					

Figure E2: First Pattern Raw Data (2 of 2)

						Second Pattern																															
	Male = 14 Female = 10					Video Start Time		Miss #1 Elapsed Time		Error Type				Miss #2 Elapsed Time		Error Type				Miss #3 Elapsed Time		Error Type				Miss #4 Elapsed Time		Error Type				Miss #5 Elapsed Time		Error Type				Video End Time	
Participant	Gender	Age	Presentation Order	Pattern A First	JPA First	Minute	Second	Minute	Second	Markings	Incorrect	Order	Orientation	Minute	Second	Markings	Incorrect	Order	Orientation	Minute	Second	Markings	Incorrect	Order	Orientation	Minute	Second	Markings	Incorrect	Order	Orientation	Minute	Second	Markings	Incorrect	Order	Orientation	Minute	Second
1	M	49	B without JPA, A with JPA			29	6																															44	37
2	F	40	A with JPA, B without JPA	1	1	22	35	38	48	1																												43	19
3	M	35	A without JPA, B with JPA	1		24	48	25	36		1			33	54	1				38	25		1			40	11	1										44	31
4	M	28	A with JPA, B without JPA	1	1	20	59	29	14	1				33	21		1			35	21	1																39	12
5	F	54	B without JPA, A with JPA			27	30	41	37	1																												41	47
6	F	30	B with JPA, A without JPA		1	23	44	27	10	1				38	16	1																						39	16
7 (DQ)	F	56	A without JPA, B with JPA																																				
8	M	30	B with JPA, A without JPA		1	31	53	35	25	1				39	28			1		46	39	1																47	39
9	F	35	B with JPA, A without JPA		1	25	4	37	30	1																												37	38
10	M	24	B without JPA, A with JPA			23	43	34	2				1	36	24																							37	30
11	F	54	A with JPA, B without JPA	1	1	19	6	19	55		1			27	47	1				34	56	1																38	30
12	M	43	A without JPA, B with JPA	1		18	0	22	52				1	26	54	1				32	54	1																36	29
13	M	49	B without JPA, A with JPA			20	5	22	48	1				29	37			1		31	38	1				31	54	1										32	0
14	M	25	A without JPA, B with JPA	1		15	55	29	59	1																												33	13
15	F	56	B with JPA, A without JPA		1	20	14	23	2	1				31	51	1				32	14	1																32	23
16	F	39	B with JPA, A without JPA		1	24	27	27	14	1				36	41	1				36	55	1																37	38
17	F	37	A with JPA, B without JPA	1	1	21	25	22	17		1			25	39			1		34	41		1			35	48	1										39	59
18	M	47	A without JPA, B with JPA	1		16	59	24	37	1				28	9		1			29	25	1																32	27
19	M	48	A with JPA, B without JPA	1	1	18	26	19	4		1			25	57	1				29	44	1				29	48		1			31	6	1				34	35
20	M	65	B without JPA, A with JPA			27	18	29	55	1				38	31	1				38	39	1																38	58
21	M	46	A without JPA, B with JPA	1		18	15	25	40	1				31	8	1																						34	48
22	M	55	A with JPA, B without JPA	1	1	17	20	25	51	1				31	55	1																						35	53
23	F	34	A without JPA, B with JPA	1		19	22	28	5	1				34	20	1																						38	21
24	M	53	B without JPA, A with JPA			24	16	26	44	1				33	5			1		35	10	1				35	35	1										35	44
25	F	58	B with JPA, A without JPA		1	28	9	30	49	1				41	49	1				42	24	1																42	32

Figure E3: Second Pattern Raw Data (1 of 2)

						Second Pattern																										
	Male = 14 Female = 10											Misses																				
Participant	Gender	Age	Presentation Order	Pattern A First	JPA First	Total Pattern Time	Miss #1 Actual Time	Miss #2 Actual Time	Miss #3 Actual Time	Miss #4 Actual Time	Miss #5 Actual Time	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	False Alarms	Mistakes	First Half Misses	Second Half Misses	Notes
1	M	49	B without JPA, A with JPA			15:31						0	0	0	0	0	0	0								0	0			
2	F	40	A with JPA, B without JPA	1	1	20:44	16:13												0	0	0		0	0	1	1	1	0	1	Only 6 errors, forgot to perform p.12 fault correctly Norman's mode error (early page turn), p.7, 28:06 elapsed on video
3	M	35	A without JPA, B with JPA	1		19:43	00:48	09:06	13:37	15:23									1	0	1	0	0	1	1	0	0	2	2	
4	M	28	A with JPA, B without JPA	1	1	18:13	08:15	12:22	14:22											0	1	0	0	1	1	0	2	1	2	Only 6 errors, forgot to perform p.2 fault correctly
5	F	54	B without JPA, A with JPA			14:17	14:07					0	0	0	0	0	0	1								0	0	0	1	
6	F	30	B with JPA, A without JPA		1	15:32	03:26	14:32				1	0	0	0	0	1	0								0	1	1	1	Only 6 errors, couldn't perform last one (1x1 eye block, p.20) because I mistakenly used it in the first pattern
7 (DQ)	F	56	A without JPA, B with JPA																											
8	M	30	B with JPA, A without JPA		1	15:46	03:32	07:35	14:46			1	0	1	0	0	1	0								0	0	2	1	Miss #1 became a catch at 41:58 Miss #3 became a catch at 47:28
9	F	35	B with JPA, A without JPA		1	12:34	12:26					0	0	0	0	0	0	1								0	0	0	1	
10	M	24	B without JPA, A with JPA			13:47	10:19	12:41				0	0	0	0	1	1	0								0	0	0	2	Miss #1 became a catch at 34:15
11	F	54	A with JPA, B without JPA	1	1	19:24	00:49	08:41	15:50										1	0	1	0	0	0	1	0	0	2	1	
12	M	43	A without JPA, B with JPA	1		18:29	04:52	08:54	14:54										0	1	1	0	0	0	1	0	0	2	1	
13	M	49	B without JPA, A with JPA			11:55	02:43	09:32	11:33	11:49		1	0	0	0	1	1	1								1	0	1	3	
14	M	25	A without JPA, B with JPA	1		17:18	14:04												0	0	0	0	0	0	1	0	0	0	1	
15	F	56	B with JPA, A without JPA		1	12:09	02:48	11:37	12:00			1	0	0	0	0	1	1								0	0	1	2	
16	F	39	B with JPA, A without JPA		1	13:11	02:47	12:14	12:28			1	0	0	0	0	1	1								0	0	1	2	
17	F	37	A with JPA, B without JPA	1	1	18:34	00:52	04:14	13:16	14:23									1	1	0	0	0	1	1	0	0	2	2	
18	M	47	A without JPA, B with JPA	1		15:28	07:38	11:10	12:26										0	0	1	0	0	1	1	1	0	1	2	
19	M	48	A with JPA, B without JPA	1	1	16:09	00:38	07:31	11:18	11:22	12:40								1	0	1	0	1	1	1	1	0	2	3	
20	M	65	B without JPA, A with JPA			11:40	02:37	11:13	11:21			1	0	0	0	0	1	1								0	0	1	2	
21	M	46	A without JPA, B with JPA	1		16:33	07:25	12:53											0	0	1	0	0	0	1	0	0	1	1	
22	M	55	A with JPA, B without JPA	1	1	18:33	08:31	14:35											0	0	1	0	0	0	1	0	0	1	1	
23	F	34	A without JPA, B with JPA	1		18:59	08:43	14:58											0	0	1	0	0	0	1	0	0	1	1	
24	M	53	B without JPA, A with JPA			11:28	02:28	08:49	10:54	11:19		1	0	0	0	1	1	1								0	0	1	3	
25	F	58	B with JPA, A without JPA		1	14:23	02:40	13:40	14:15			1	0	0	0	0	1	1								0	0	1	2	
												8	0	1	0	3	9	8	4	2	9	0	1	5	12					

Figure E4: Second Pattern Raw Data (2 of 2)

Appendix F Fault Types

Fault #1, detailed in Chapter 3, is reproduced here:

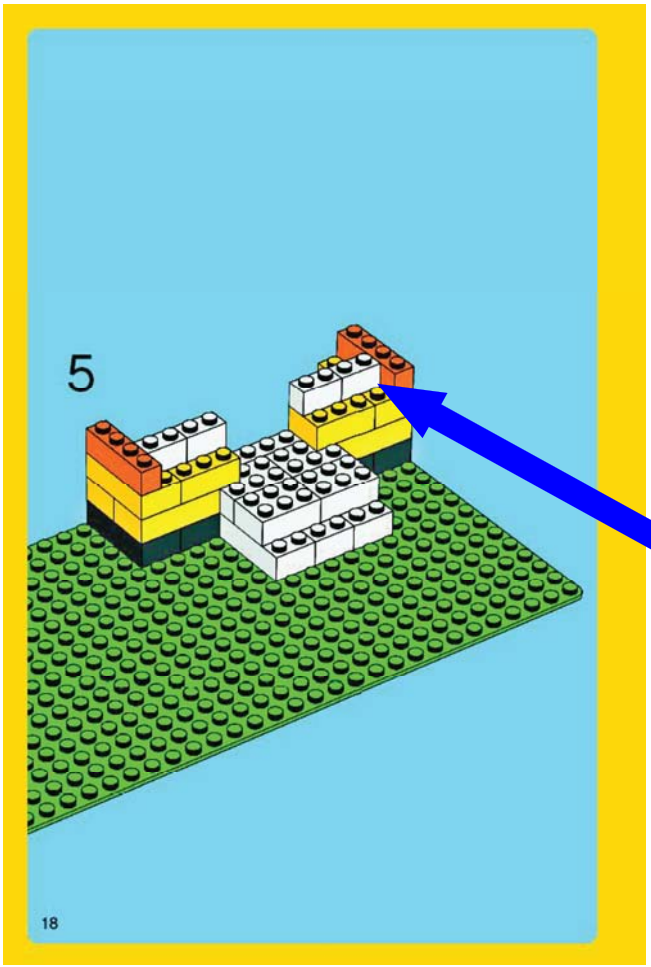


Figure F1: Sample Instructions Page, Depicting Fault #1 (Fault Type 1)

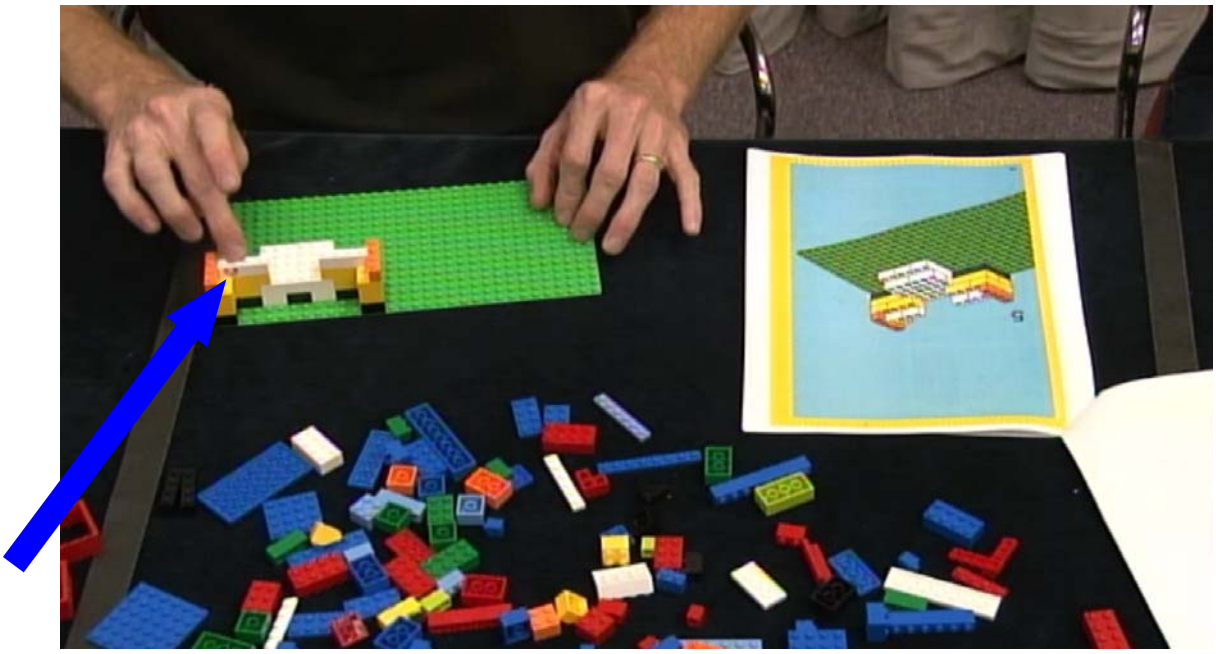


Figure F2: Fault #1 Occurring in an Experimental Trial

Figure F3 shows a page from the assembly instructions for fault type 2 (Fault #8), which is where an incorrect assembly piece is used. The arrow points to the black 2 x 4 piece that is replaced by two 1 x 4 pieces in the experimental trial, one of which is shown in Figure F4.

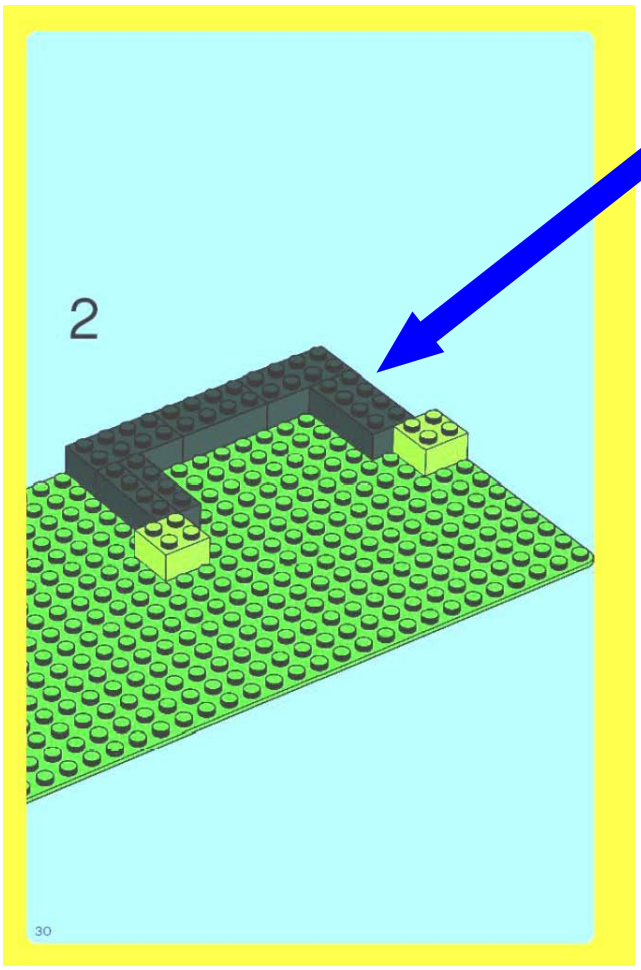


Figure F3: Sample Instructions Page, Depicting Fault #8 (Fault Type 2)

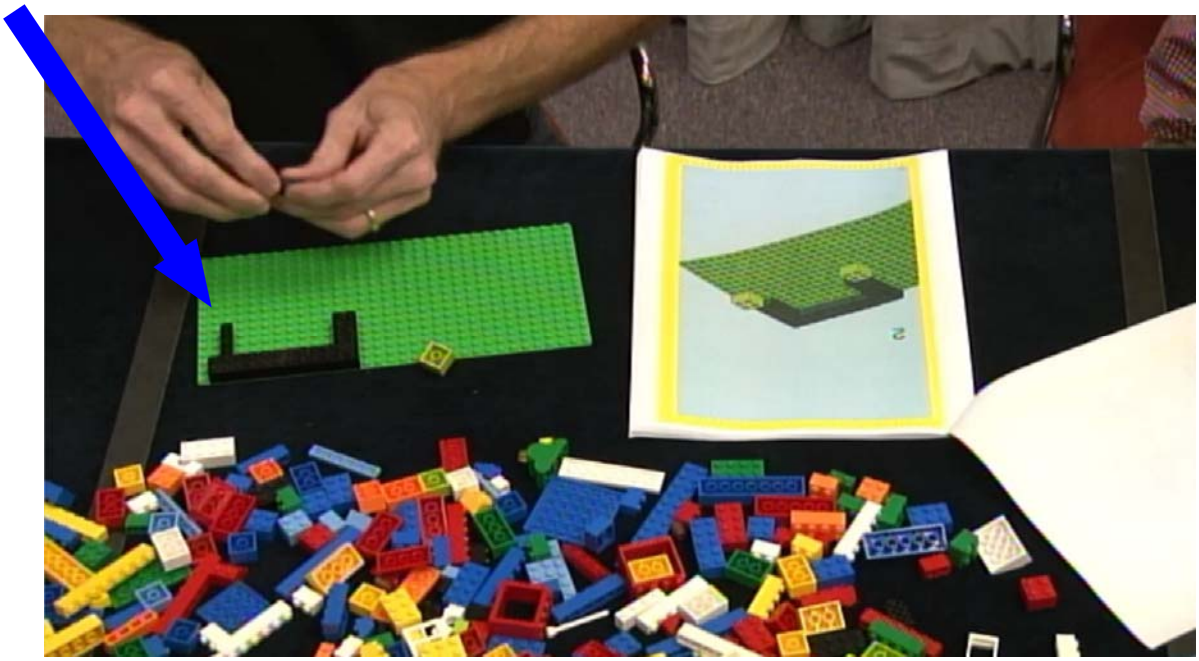


Figure F4: Fault #8 Occurring in an Experimental Trial

Figure F5 depicts fault type 3 (also Fault #3) whereby the correct pieces are used but they are not in the correct order. The arrow shows the correct orientation where the 2 x 2 piece is in the rear corner of the assembly. Figure F6, however, indicates the fault whereby the 2 x 4 piece is moved such that it appears in the rear corner instead.

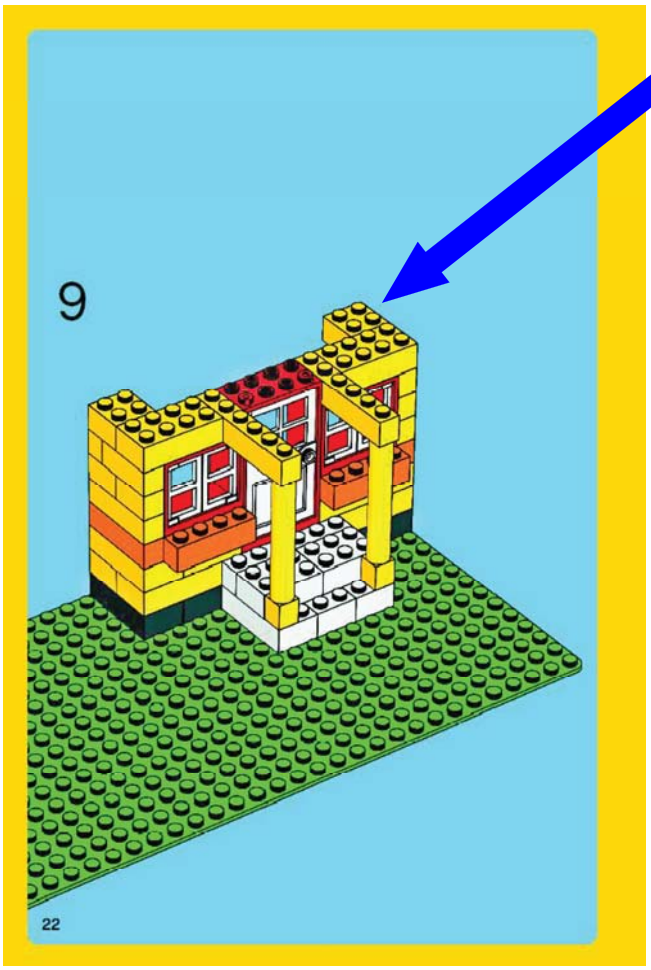


Figure F5: Sample Instructions Page, Depicting Fault #3 (Fault Type 3)

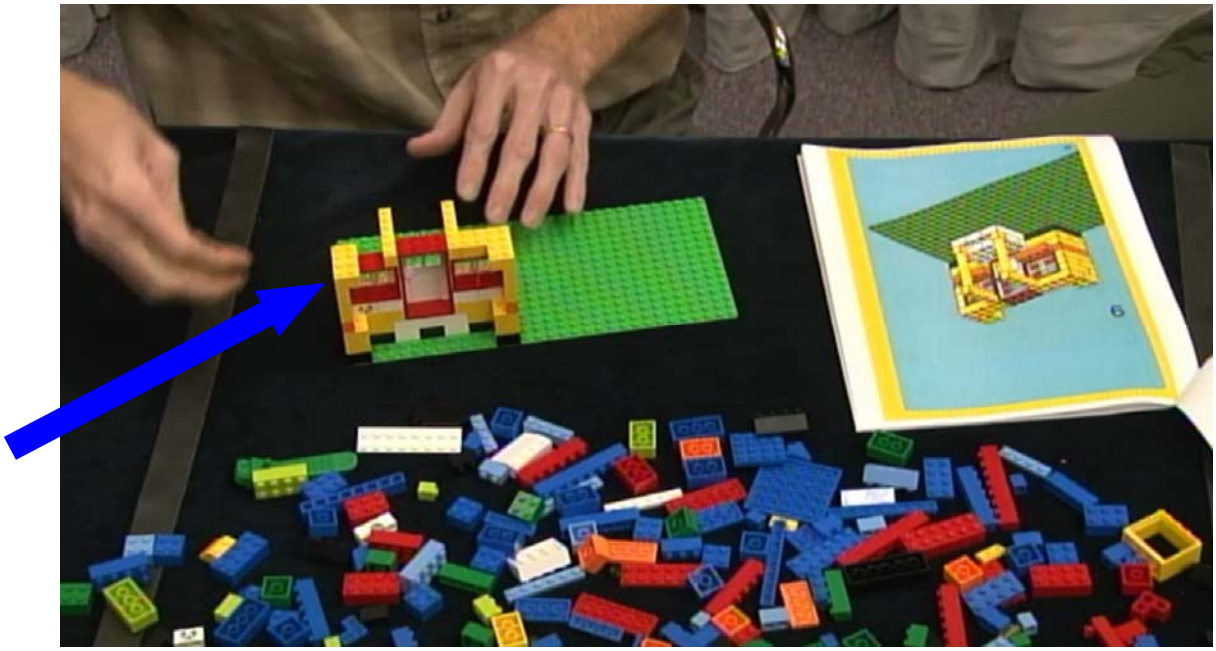


Figure F6: Fault #3 Occurring in an Experimental Trial

Figure F7 shows a red window in the assembly that is correctly installed. An example of fault type 4, which uses a correct part but in the wrong orientation, is seen in Figure F8 where the window is installed backwards.

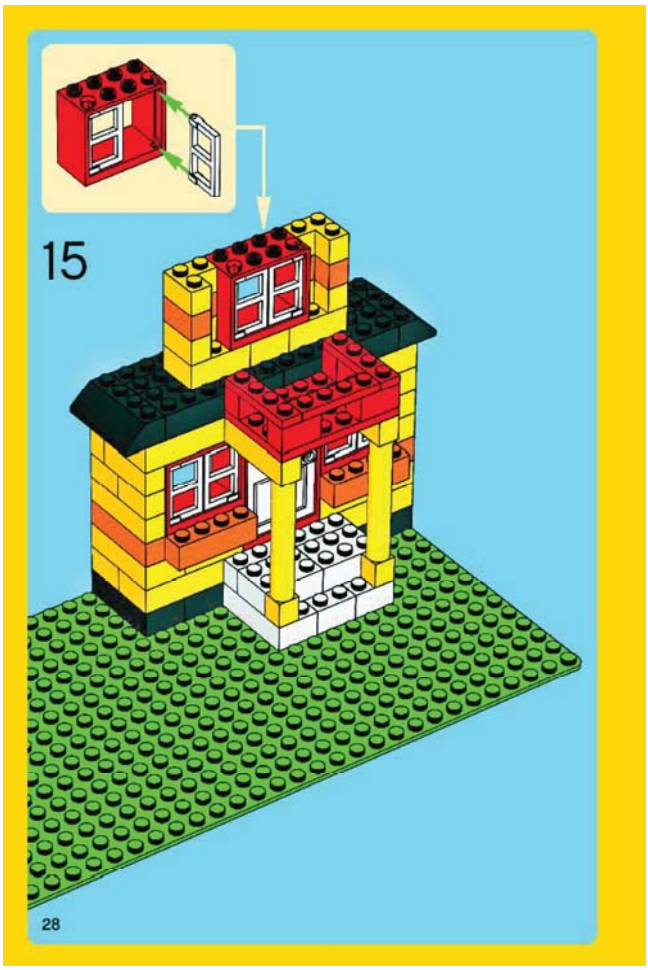


Figure F7: Sample Instructions Page, Depicting Fault #5 (Fault Type 4)

Figure F8: Fault #5 Occurring in an Experimental Trial

Appendix G Minitab Results

The following data are a direct output from Minitab analyses. Note that there were two separate series of calculations run on 12/9/2013 and 1/2/2014.

12/9/2013 3:35:45 PM

Results for: Pattern A

Descriptive Statistics: Result

Variable	ErrorNumber	Sum
Result	1	19.0000
	2	0.000000
	3	1.0000
	4	0.000000
	5	7.0000
	6	18.0000
	7	18.0000

7 column(s) excluded because they have the wrong number of rows.
Subset worksheet Subset of Pattern A created.

Results for: Pattern A wo S1 E2 E4

Binary Logistic Regression: Result versus Sequence, ErrorNumber

Link Function: Logit

Response Information

Variable	Value	Count	
Result	1	63	(Event)
	0	52	
	Total	115	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	2.97386	0.850309	3.50	0.000			
Sequence							
BJA	-1.20882	0.821704	-1.47	0.141	0.30	0.06	1.49
JAB	-2.06992	0.796186	-2.60	0.009	0.13	0.03	0.60
JBA	-1.63687	0.792043	-2.07	0.039	0.19	0.04	0.92
ErrorNumber							
3	-5.13880	1.23813	-4.15	0.000	0.01	0.00	0.07
5	-2.67489	0.773096	-3.46	0.001	0.07	0.02	0.31
6	-0.295419	0.771207	-0.38	0.702	0.74	0.16	3.37
7	-0.295419	0.771207	-0.38	0.702	0.74	0.16	3.37

```
Log-Likelihood = -48.714
Test that all slopes are zero: G = 60.943, DF = 7, P-Value = 0.000


Goodness-of-Fit Tests

Method           Chi-Square  DF      P
Pearson             16.2009  12  0.182
Deviance            15.0132  12  0.241
Hosmer-Lemeshow      7.8827   8  0.445


Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

                                Group
Value     1     2    3    4    5    6    7     8     9   10  Total
1
  Obs     1     0    1    6    7   10    9    11    12    6     63
  Exp   0.2   0.8  2.1  4.9  7.8  8.7  9.2  12.4  11.2  5.7
0
  Obs    11    11   11    5    5    2    3     4     0    0     52
  Exp  11.8  10.2  9.9  6.1  4.2  3.3  2.8   2.6   0.8  0.3
Total    12    11   12   11   12   12   12    15    12    6    115


Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs       Number  Percent  Summary Measures
Concordant    2814     85.9  Somers' D              0.75
Discordant     343     10.5  Goodman-Kruskal Gamma  0.78
Ties           119      3.6  Kendall's Tau-a        0.38
Total         3276    100.0

13 column(s) excluded because they have the wrong number of rows.
Subset worksheet Subset of Pattern A created.

13 column(s) excluded because they have the wrong number of rows.
Subset worksheet Subset of Pattern A created.
```

Results for: Pattern A wo E2 E4

Binary Logistic Regression: Result versus Sequence, ErrorNumber

```
Link Function: Logit


Response Information

Variable  Value  Count
Result    1         63  (Event)
          0         57
          Total    120


Logistic Regression Table
```

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	2.84529	0.809862	3.51	0.000			
Sequence							
BJA	-1.79208	0.776314	-2.31	0.021	0.17	0.04	0.76
JAB	-1.99867	0.778398	-2.57	0.010	0.14	0.03	0.62
JBA	-1.57763	0.774927	-2.04	0.042	0.21	0.05	0.94
ErrorNumber							
3	-4.96722	1.21817	-4.08	0.000	0.01	0.00	0.08
5	-2.49396	0.731281	-3.41	0.001	0.08	0.02	0.35
6	-0.250882	0.709869	-0.35	0.724	0.78	0.19	3.13
7	-0.250882	0.709869	-0.35	0.724	0.78	0.19	3.13

Log-Likelihood = -53.394
Test that all slopes are zero: G = 59.267, DF = 7, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	16.4607	12	0.171
Deviance	15.9309	12	0.194
Hosmer-Lemeshow	7.2051	8	0.515

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

	Group										
Value	1	2	3	4	5	6	7	8	9	10	Total
1											
Obs	0	1	4	3	7	8	10	7	11	12	63
Exp	0.2	0.8	2.1	4.9	7.7	8.3	8.6	8.9	10.3	11.3	
0											
Obs	12	11	8	9	5	4	2	5	1	0	57
Exp	11.8	11.2	9.9	7.1	4.3	3.7	3.4	3.1	1.7	0.7	
Total	12	12	12	12	12	12	12	12	12	12	120

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	3030	84.4	Somers' D	0.73
Discordant	420	11.7	Goodman-Kruskal Gamma	0.76
Ties	141	3.9	Kendall's Tau-a	0.37
Total	3591	100.0		

Binary Logistic Regression: Result versus Sequence, ErrorNumber

Link Function: Logit

Response Information

Variable	Value	Count	
Result	1	63	(Event)
	0	57	
	Total	120	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	2.84529	0.809862	3.51	0.000			
Sequence							
BJA	-1.79208	0.776314	-2.31	0.021	0.17	0.04	0.76
JAB	-1.99867	0.778398	-2.57	0.010	0.14	0.03	0.62
JBA	-1.57763	0.774927	-2.04	0.042	0.21	0.05	0.94
ErrorNumber							
3	-4.96722	1.21817	-4.08	0.000	0.01	0.00	0.08
5	-2.49396	0.731281	-3.41	0.001	0.08	0.02	0.35
6	-0.250882	0.709869	-0.35	0.724	0.78	0.19	3.13
7	-0.250882	0.709869	-0.35	0.724	0.78	0.19	3.13

Log-Likelihood = -53.394
Test that all slopes are zero: G = 59.267, DF = 7, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	16.4607	12	0.171
Deviance	15.9309	12	0.194
Hosmer-Lemeshow	7.2051	8	0.515

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

	Group										
Value	1	2	3	4	5	6	7	8	9	10	Total
1											
Obs	0	1	4	3	7	8	10	7	11	12	63
Exp	0.2	0.8	2.1	4.9	7.7	8.3	8.6	8.9	10.3	11.3	
0											
Obs	12	11	8	9	5	4	2	5	1	0	57
Exp	11.8	11.2	9.9	7.1	4.3	3.7	3.4	3.1	1.7	0.7	
Total	12	12	12	12	12	12	12	12	12	12	120

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	3030	84.4	Somers' D	0.73
Discordant	420	11.7	Goodman-Kruskal Gamma	0.76
Ties	141	3.9	Kendall's Tau-a	0.37
Total	3591	100.0		

Binary Logistic Regression: Result versus Sequence, ErrorNumber

Link Function: Logit

Response Information

```
Variable  Value  Count
Result    1         63  (Event)
          0         57
          Total    120


Logistic Regression Table

                                                    Odds     95% CI
Predictor         Coef   SE Coef      Z      P  Ratio  Lower  Upper
Constant       2.84529  0.809862   3.51  0.000
Sequence
 BJA          -1.79208  0.776314  -2.31  0.021   0.17   0.04   0.76
 JAB          -1.99867  0.778398  -2.57  0.010   0.14   0.03   0.62
 JBA          -1.57763  0.774927  -2.04  0.042   0.21   0.05   0.94
ErrorNumber
 3            -4.96722   1.21817  -4.08  0.000   0.01   0.00   0.08
 5            -2.49396  0.731281  -3.41  0.001   0.08   0.02   0.35
 6           -0.250882  0.709869  -0.35  0.724   0.78   0.19   3.13
 7           -0.250882  0.709869  -0.35  0.724   0.78   0.19   3.13


Log-Likelihood = -53.394
Test that all slopes are zero: G = 59.267, DF = 7, P-Value = 0.000


Goodness-of-Fit Tests

Method           Chi-Square  DF      P
Pearson             16.4607  12  0.171
Deviance            15.9309  12  0.194
Hosmer-Lemeshow      7.2051   8  0.515


Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

                              Group
Value     1     2    3    4    5    6    7    8     9    10  Total
1
  Obs     0     1    4    3    7    8   10    7    11    12     63
  Exp   0.2   0.8  2.1  4.9  7.7  8.3  8.6  8.9  10.3  11.3
0
  Obs    12    11    8    9    5    4    2    5     1     0     57
  Exp  11.8  11.2  9.9  7.1  4.3  3.7  3.4  3.1   1.7   0.7
Total    12    12   12   12   12   12   12   12    12    12    120


Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs       Number  Percent  Summary Measures
Concordant    3030     84.4  Somers' D              0.73
Discordant     420     11.7  Goodman-Kruskal Gamma  0.76
Ties           141      3.9  Kendall's Tau-a        0.37
Total         3591    100.0
```

Results for: Pattern B

Descriptive Statistics: Result

Variable	ErrorNumber	Sum
Result	1	3.000
	2	1.0000
	3	10.000
	4	0.000000
	5	3.000
	6	2.000
	7	10.000

Descriptive Statistics: Result

Variable	ErrorNumber	Sum
Result	1	5.0000
	2	10.000
	3	10.000
	4	1.0000
	5	8.0000
	6	14.000
	7	14.000

Descriptive Statistics: Result

Variable	ErrorNumber	Sum
Result	1	7.000
	2	3.0000
	3	19.0000
	4	0.000000
	5	4.0000
	6	7.0000
	7	22.0000

13 column(s) excluded because they have the wrong number of rows.
Subset worksheet Pattern B wo E4 created.

Results for: Pattern B wo E4

Binary Logistic Regression: Result versus Sequence, ErrorNumber

Link Function: Logit

Response Information

Variable	Value	Count	
Result	1	62	(Event)
	0	80	
	Total	142	

* NOTE * 142 cases were used
* NOTE * 2 cases contained missing values

Logistic Regression Table

				Odds	95% CI

Predictor	Coef	SE Coef	Z	P	Ratio	Lower	Upper
Constant	-0.892876	0.593447	-1.50	0.132			
Sequence							
BJA	-0.350493	0.636774	-0.55	0.582	0.70	0.20	2.45
JAB	0.614972	0.610142	1.01	0.313	1.85	0.56	6.12
JBA	0.186810	0.611754	0.31	0.760	1.21	0.36	4.00
ErrorNumber							
2	-1.21127	0.776262	-1.56	0.119	0.30	0.07	1.36
3	2.15010	0.691037	3.11	0.002	8.59	2.22	33.26
5	-0.869622	0.721889	-1.20	0.228	0.42	0.10	1.73
6	-0.132053	0.649534	-0.20	0.839	0.88	0.25	3.13
7	3.22739	0.879472	3.67	0.000	25.21	4.50	141.33

Log-Likelihood = -66.020
Test that all slopes are zero: G = 62.525, DF = 8, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	15.7178	15	0.401
Deviance	17.4991	15	0.290
Hosmer-Lemeshow	6.9396	6	0.326

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

				Group					
Value	1	2	3	4	5	6	7	8	Total
1									
Obs	4	1	3	4	6	12	14	18	62
Exp	1.8	2.7	3.4	4.8	6.2	11.1	15.3	16.7	
0									
Obs	14	17	14	14	12	5	4	0	80
Exp	16.2	15.3	13.6	13.2	11.8	5.9	2.7	1.3	
Total	18	18	17	18	18	17	18	18	142

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	4131	83.3	Somers' D	0.69
Discordant	703	14.2	Goodman-Kruskal Gamma	0.71
Ties	126	2.5	Kendall's Tau-a	0.34
Total	4960	100.0		

1/2/2014 10:11:20 AM

Welcome to Minitab, press F1 for help.
Retrieving project from file:
'C:\PROJECTS\ACTIVE\FOSSHAGE\REDO_2\MINITAB.MPJ'

Results for: Pattern A wo E2 E4

Binary Logistic Regression: RR versus Sequence, ErrorNumber

Link Function: Logit

Response Information

Variable	Value	Count	
RR	1	57	(Event)
	0	63	
	Total	120	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-2.84529	0.809862	-3.51	0.000			
Sequence							
BJA	1.79208	0.776314	2.31	0.021	6.00	1.31	27.48
JAB	1.99867	0.778398	2.57	0.010	7.38	1.60	33.93
JBA	1.57763	0.774927	2.04	0.042	4.84	1.06	22.12
ErrorNumber							
3	4.96722	1.21817	4.08	0.000	143.63	13.19	1563.68
5	2.49396	0.731281	3.41	0.001	12.11	2.89	50.77
6	0.250882	0.709869	0.35	0.724	1.29	0.32	5.17
7	0.250882	0.709869	0.35	0.724	1.29	0.32	5.17

Log-Likelihood = -53.394
Test that all slopes are zero: G = 59.267, DF = 7, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	16.4607	12	0.171
Deviance	15.9309	12	0.194
Hosmer-Lemeshow	5.3029	8	0.725

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

	Group										
Value	1	2	3	4	5	6	7	8	9	10	Total
1											
Obs	0	1	4	3	4	5	9	8	11	12	57
Exp	0.7	1.7	3.1	3.4	3.7	4.3	7.1	9.9	11.2	11.8	
0											
Obs	12	11	8	9	8	7	3	4	1	0	63
Exp	11.3	10.3	8.9	8.6	8.3	7.7	4.9	2.1	0.8	0.2	
Total	12	12	12	12	12	12	12	12	12	12	120

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	3030	84.4	Somers' D	0.73
Discordant	420	11.7	Goodman-Kruskal Gamma	0.76
Ties	141	3.9	Kendall's Tau-a	0.37

```
Total         3591    100.0
```

Results for: Pattern B wo E4

Binary Logistic Regression: RR versus Sequence, ErrorNumber

```
Link Function: Logit


Response Information

Variable  Value  Count
RR        1         80  (Event)
          0         62
          Total    142

* NOTE * 142 cases were used
* NOTE * 2 cases contained missing values


Logistic Regression Table

                                                    Odds     95% CI
Predictor         Coef   SE Coef      Z      P  Ratio  Lower  Upper
Constant      0.892876  0.593447   1.50  0.132
Sequence
 BJA          0.350493  0.636774   0.55  0.582   1.42   0.41   4.95
 JAB         -0.614972  0.610142  -1.01  0.313   0.54   0.16   1.79
 JBA         -0.186810  0.611754  -0.31  0.760   0.83   0.25   2.75
ErrorNumber
 2            1.21127  0.776262   1.56  0.119   3.36   0.73  15.37
 3           -2.15010  0.691037  -3.11  0.002   0.12   0.03   0.45
 5            0.869622  0.721889   1.20  0.228   2.39   0.58   9.82
 6            0.132053  0.649534   0.20  0.839   1.14   0.32   4.08
 7           -3.22739  0.879472  -3.67  0.000   0.04   0.01   0.22


Log-Likelihood = -66.020
Test that all slopes are zero: G = 62.525, DF = 8, P-Value = 0.000


Goodness-of-Fit Tests

Method           Chi-Square  DF      P
Pearson             15.7178  15  0.401
Deviance            17.4991  15  0.290
Hosmer-Lemeshow      6.9396   6  0.326


Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

                          Group
Value     1     2     3     4     5     6     7     8  Total
1
  Obs     0     4     5    12    14    14    17    14     80
  Exp   1.3   2.7   5.9  11.8  13.2  13.6  15.3  16.2
0
  Obs    18    14    12     6     4     3     1     4     62
  Exp  16.7  15.3  11.1   6.2   4.8   3.4   2.7   1.8
```

```
Total    18    18    17    18    18    17    18    18    142


Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs       Number  Percent  Summary Measures
Concordant    4131     83.3  Somers' D              0.69
Discordant     703     14.2  Goodman-Kruskal Gamma  0.71
Ties           126      2.5  Kendall's Tau-a        0.34
Total         4960    100.0
```

Results for: Pattern A wo E2 E4

Binary Logistic Regression: RR versus Sequence, ErrorNumber

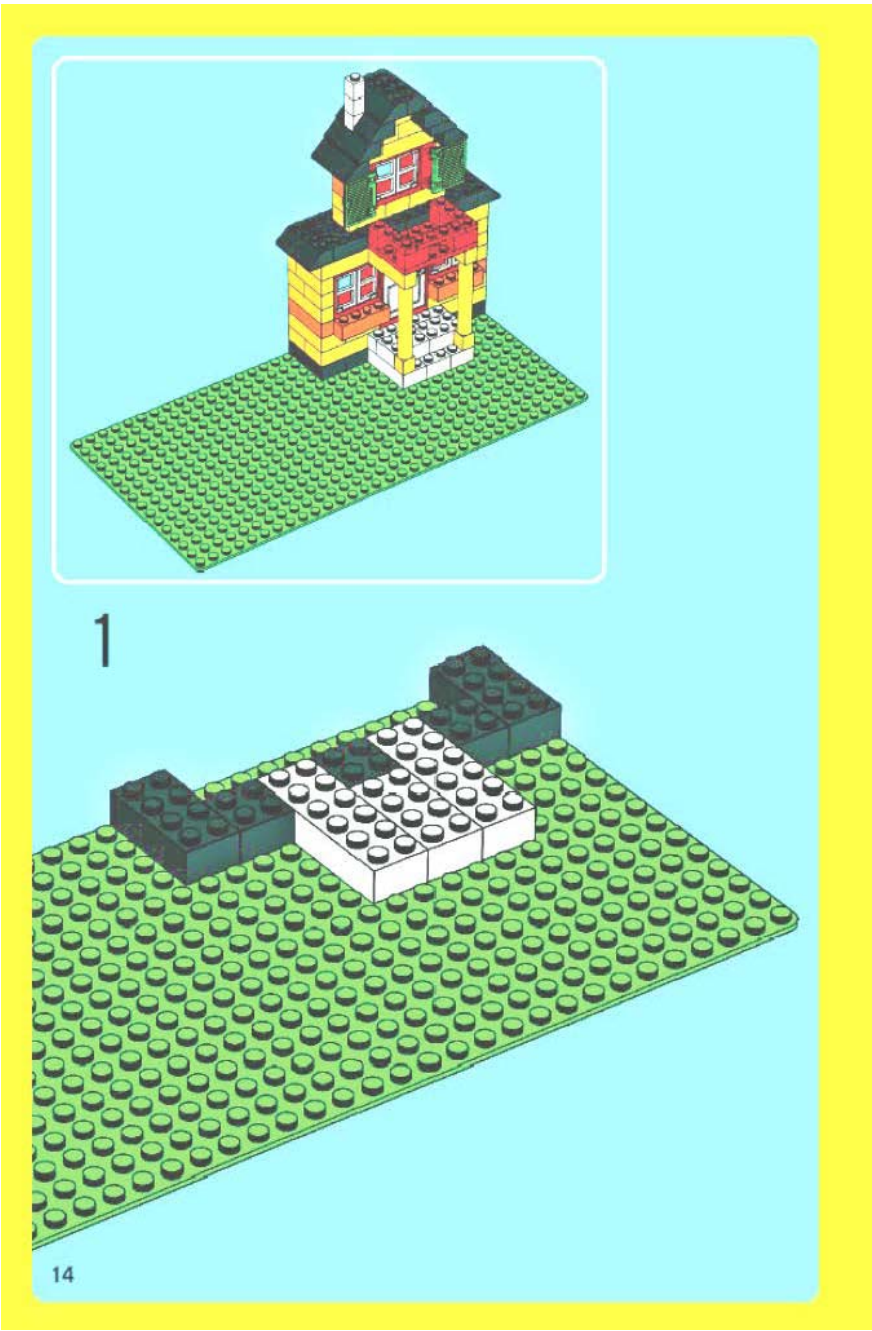
```
Link Function: Logit


Response Information

Variable  Value  Count
RR        1         57  (Event)
          0         63
          Total    120


Logistic Regression Table

                                                                  95% CI
Predictor        Coef   SE Coef      Z      P  Odds Ratio  Lower    Upper
Constant     -2.84529  0.809862  -3.51  0.000
Sequence
 BJA          1.79208  0.776314   2.31  0.021        6.00   1.31    27.48
 JAB          1.99867  0.778398   2.57  0.010        7.38   1.60    33.93
 JBA          1.57763  0.774927   2.04  0.042        4.84   1.06    22.12
ErrorNumber
 3            4.96722   1.21817   4.08  0.000      143.63  13.19  1563.68
 5            2.49396  0.731281   3.41  0.001       12.11   2.89    50.77
 6           0.250882  0.709869   0.35  0.724        1.29   0.32     5.17
 7           0.250882  0.709869   0.35  0.724        1.29   0.32     5.17


Log-Likelihood = -53.394
Test that all slopes are zero: G = 59.267, DF = 7, P-Value = 0.000


Goodness-of-Fit Tests

Method           Chi-Square  DF      P
Pearson             16.4607  12  0.171
Deviance            15.9309  12  0.194
Hosmer-Lemeshow      5.3029   8  0.725


Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

                            Group
Value     1     2    3    4    5    6    7    8     9    10  Total
```

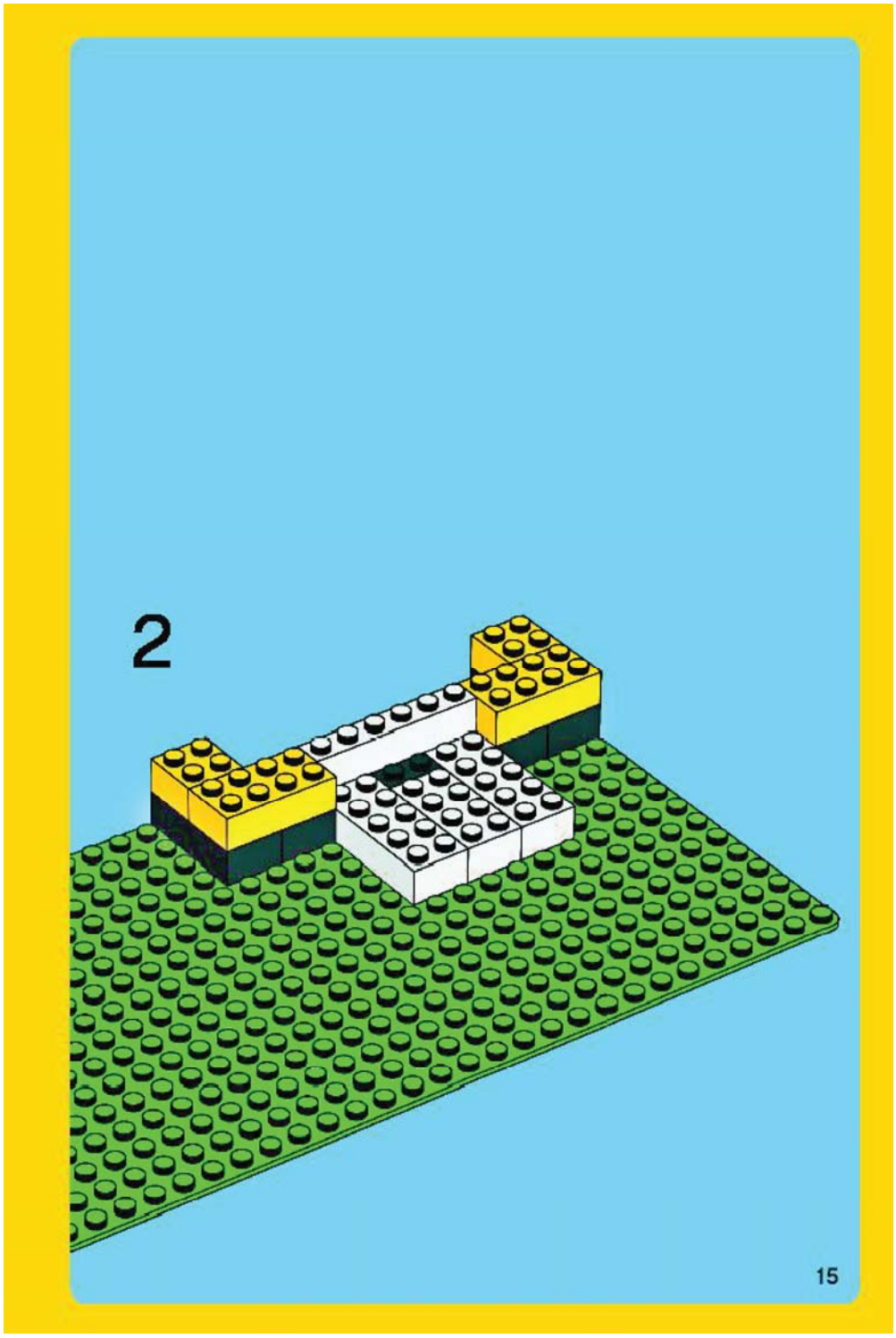
1											
	Obs	0	1	4	3	4	5	9	8	11	12	57
	Exp	0.7	1.7	3.1	3.4	3.7	4.3	7.1	9.9	11.2	11.8	
0												
	Obs	12	11	8	9	8	7	3	4	1	0	63
	Exp	11.3	10.3	8.9	8.6	8.3	7.7	4.9	2.1	0.8	0.2	
Total		12	12	12	12	12	12	12	12	12	12	120

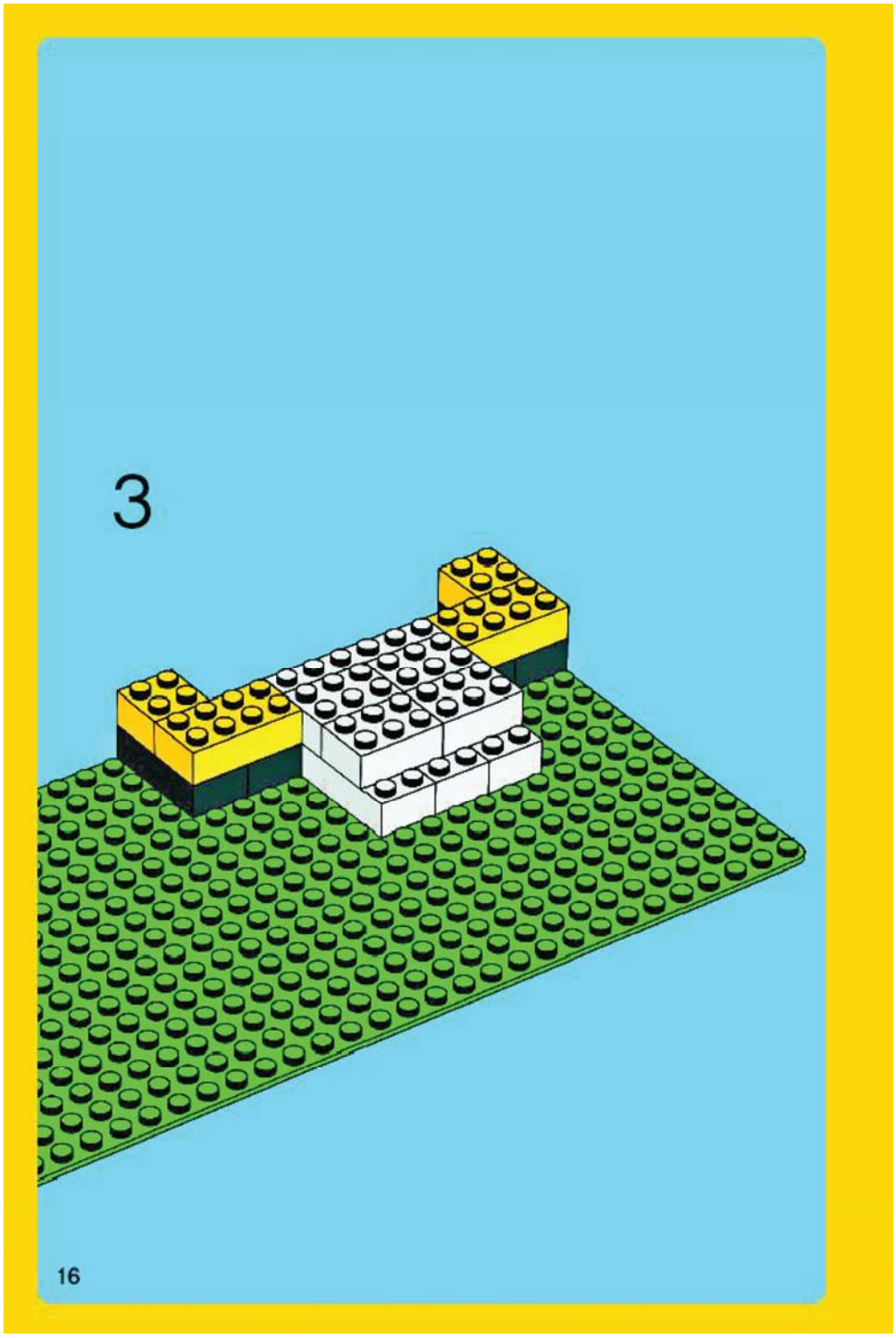
Measures of Association:
(Between the Response Variable and Predicted Probabilities)

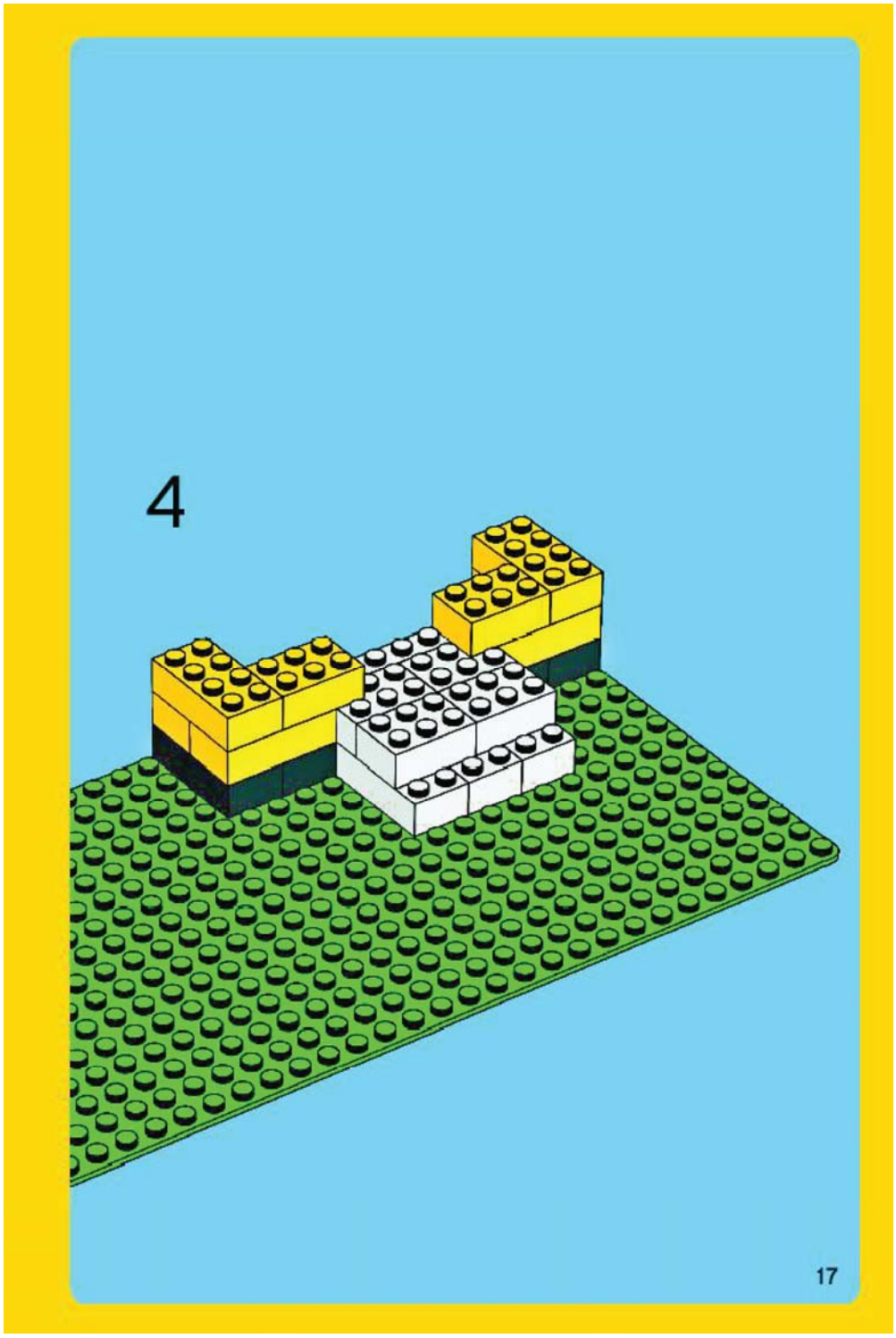
Pairs	Number	Percent	Summary Measures	
Concordant	3030	84.4	Somers' D	0.73
Discordant	420	11.7	Goodman-Kruskal Gamma	0.76
Ties	141	3.9	Kendall's Tau-a	0.37
Total	3591	100.0		

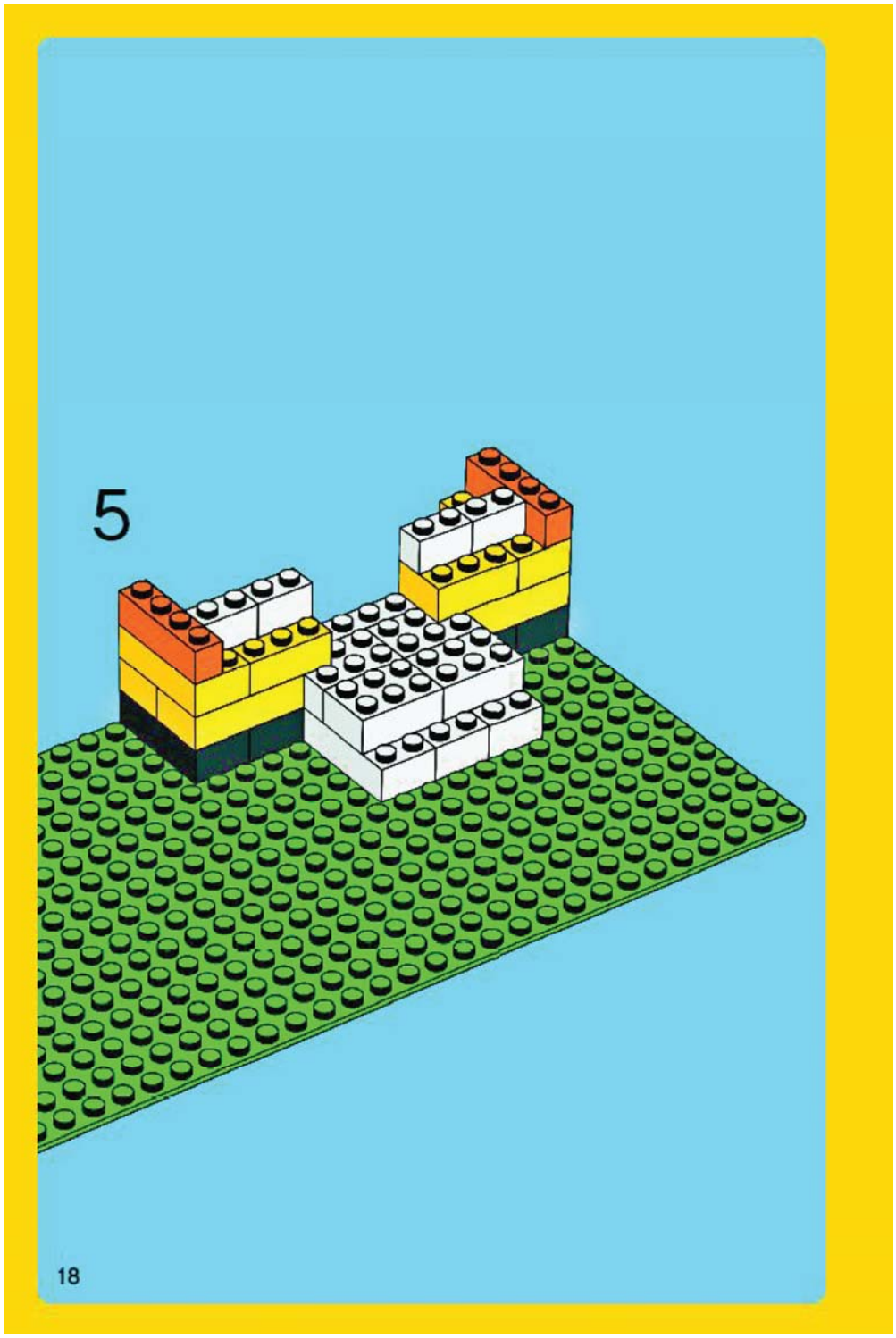
Appendix H Assembly Instructions from Pattern A (Lego™ Kit 5508)

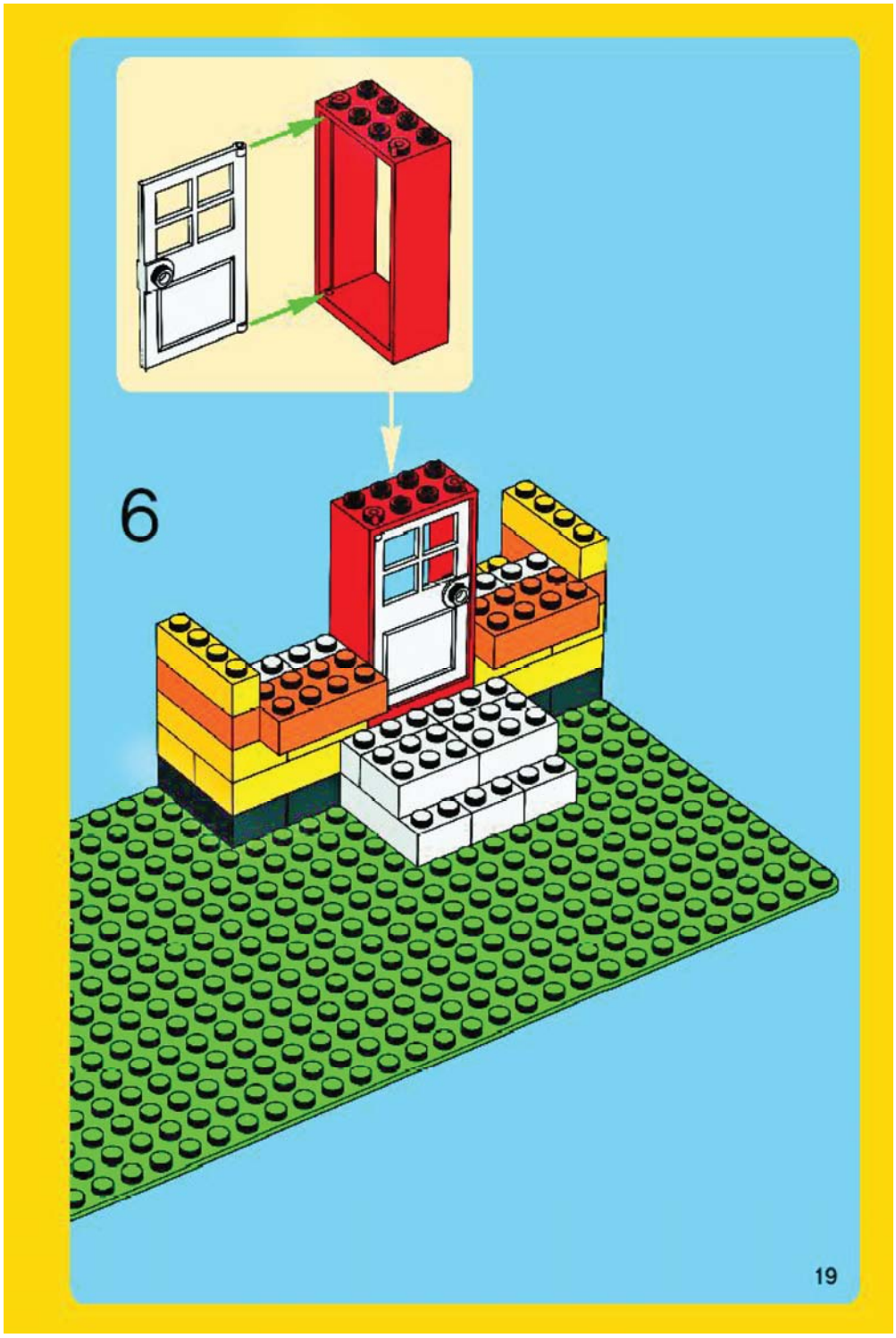


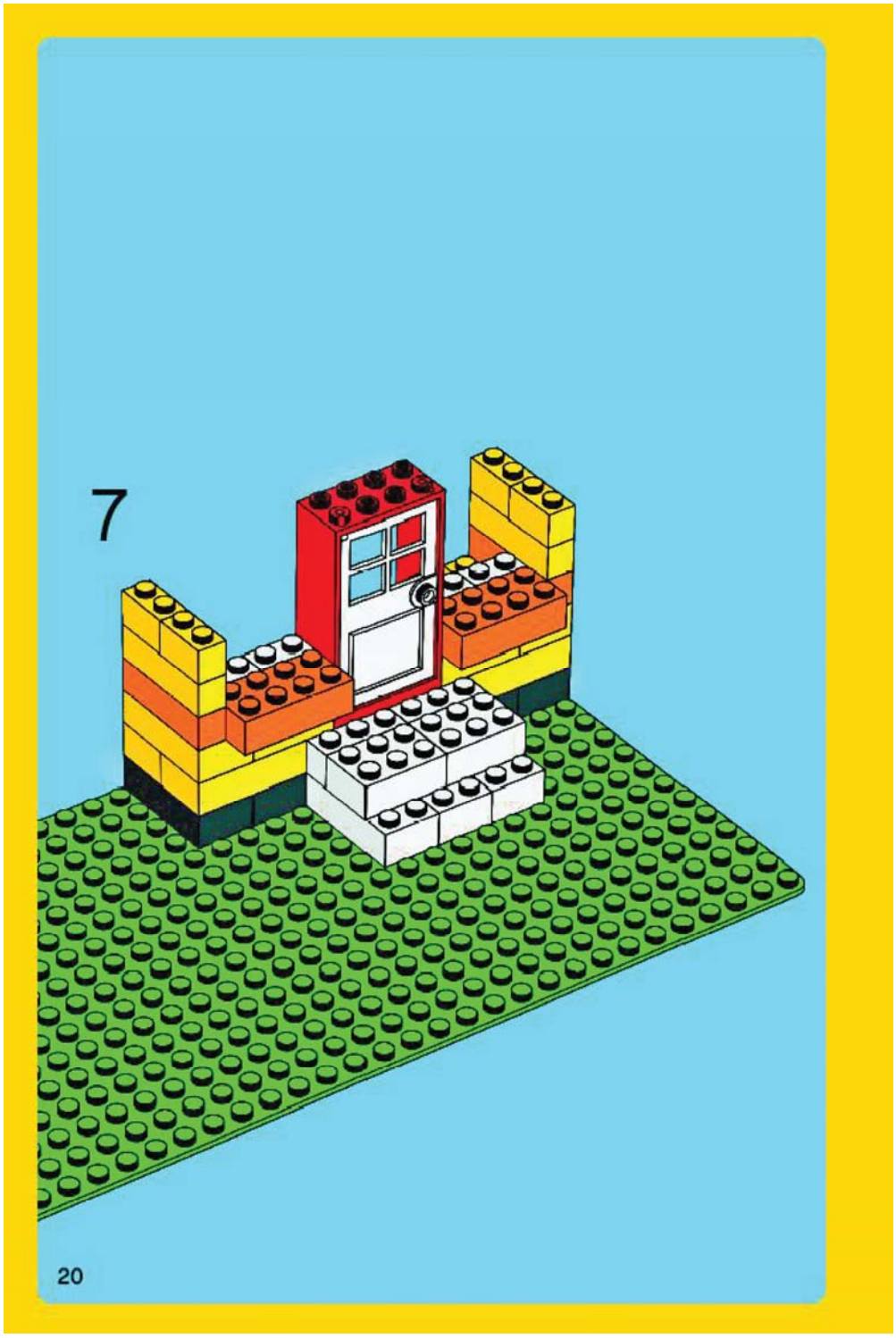

2
15

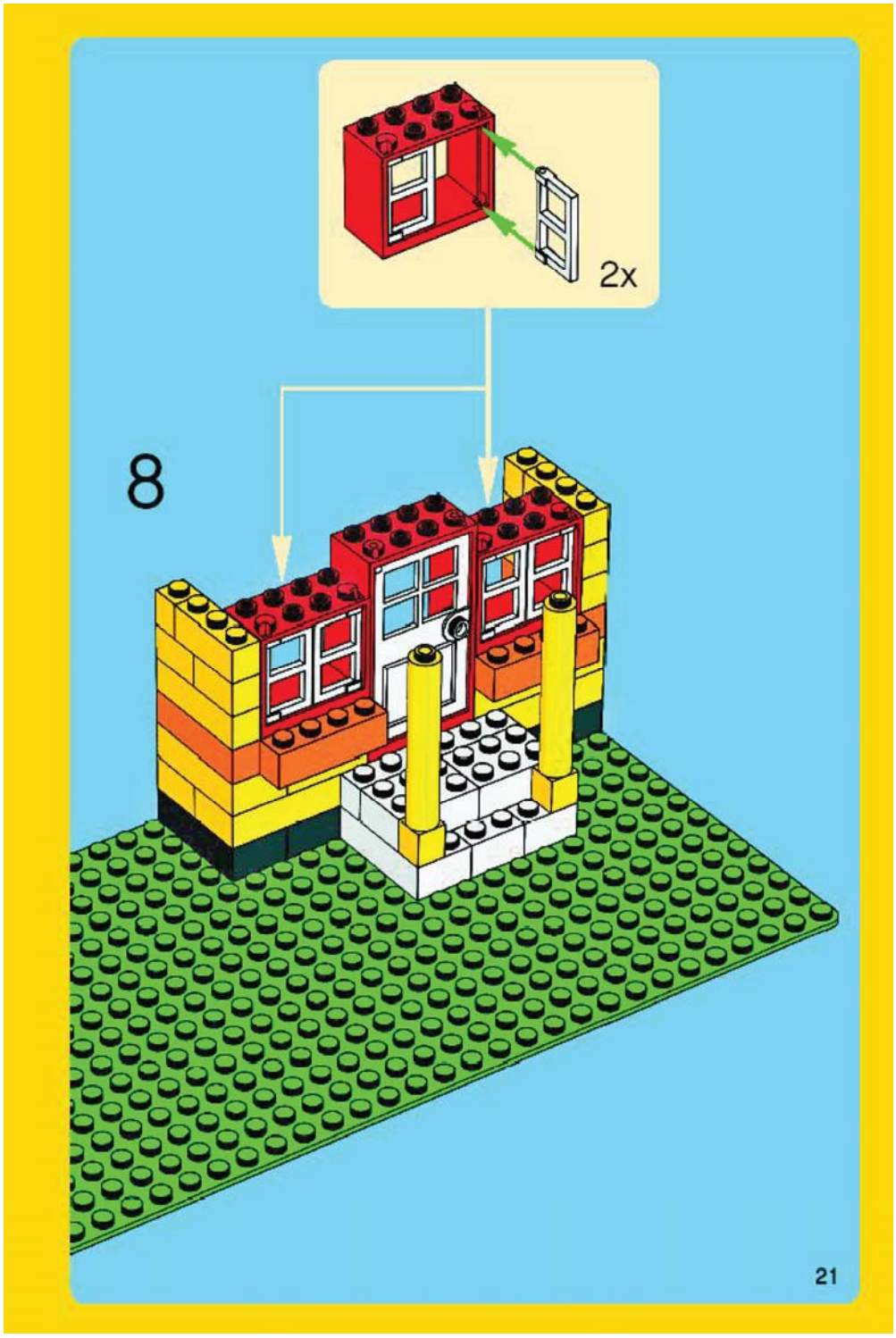

3
16

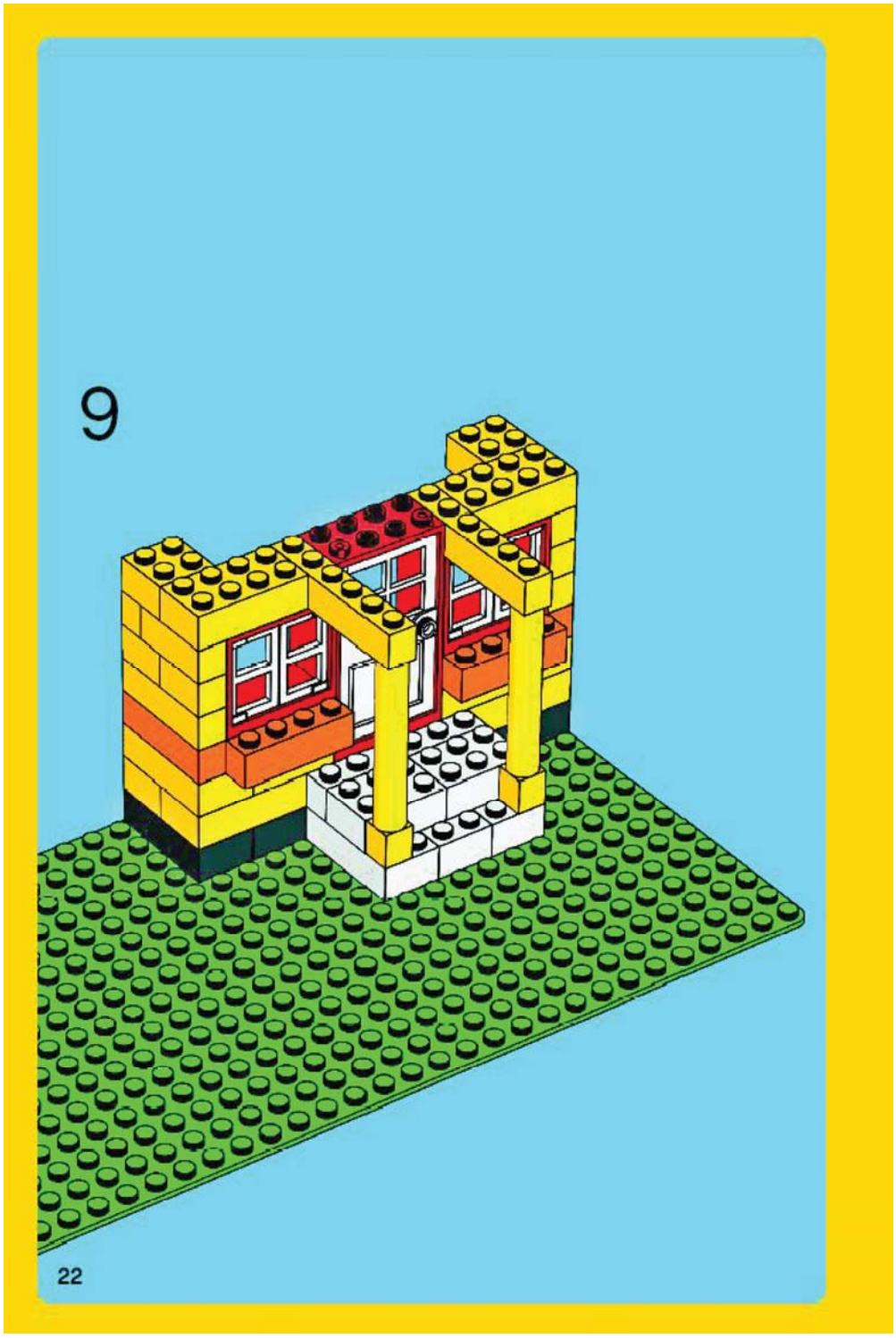

4
17

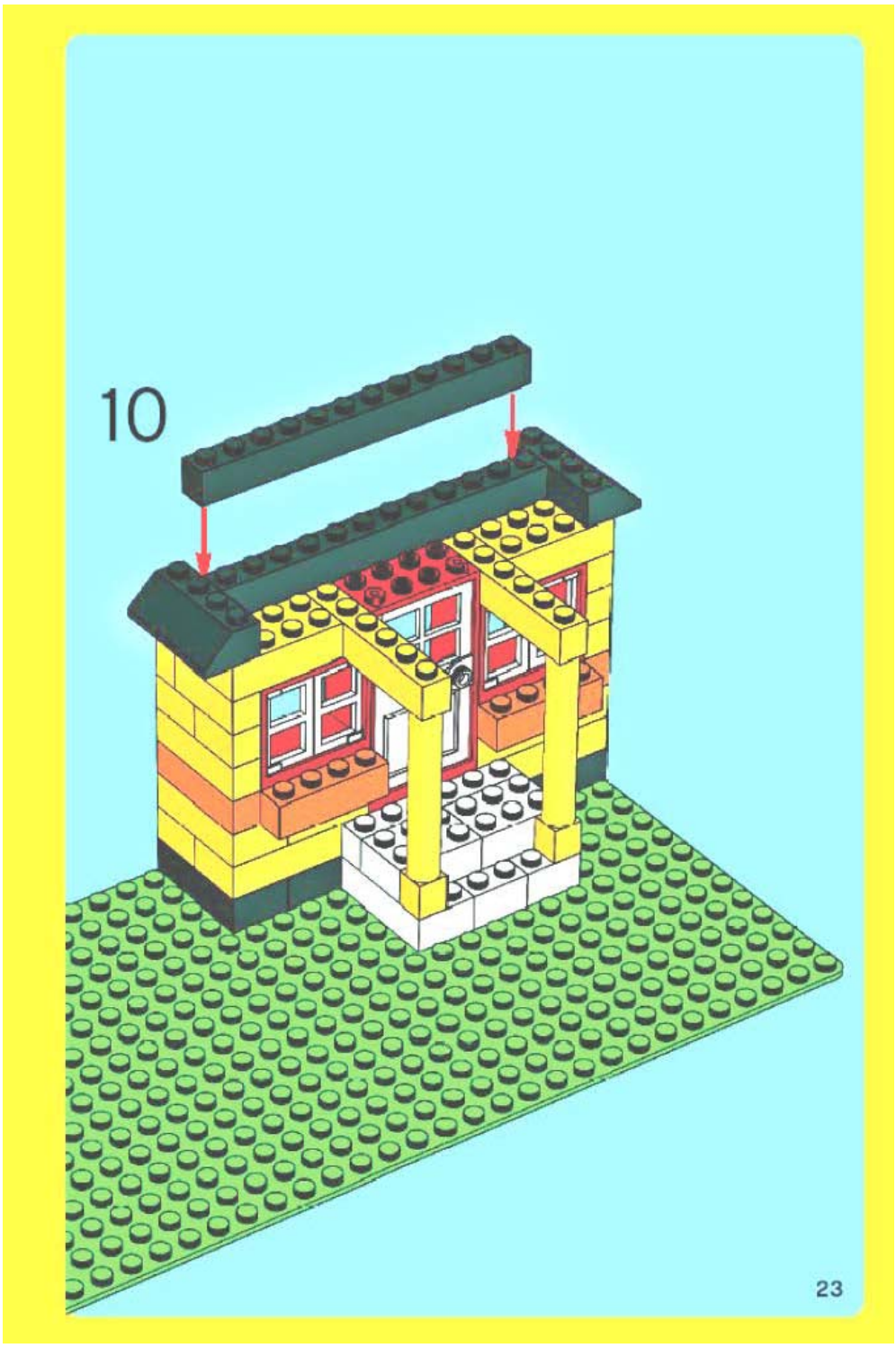

5
18

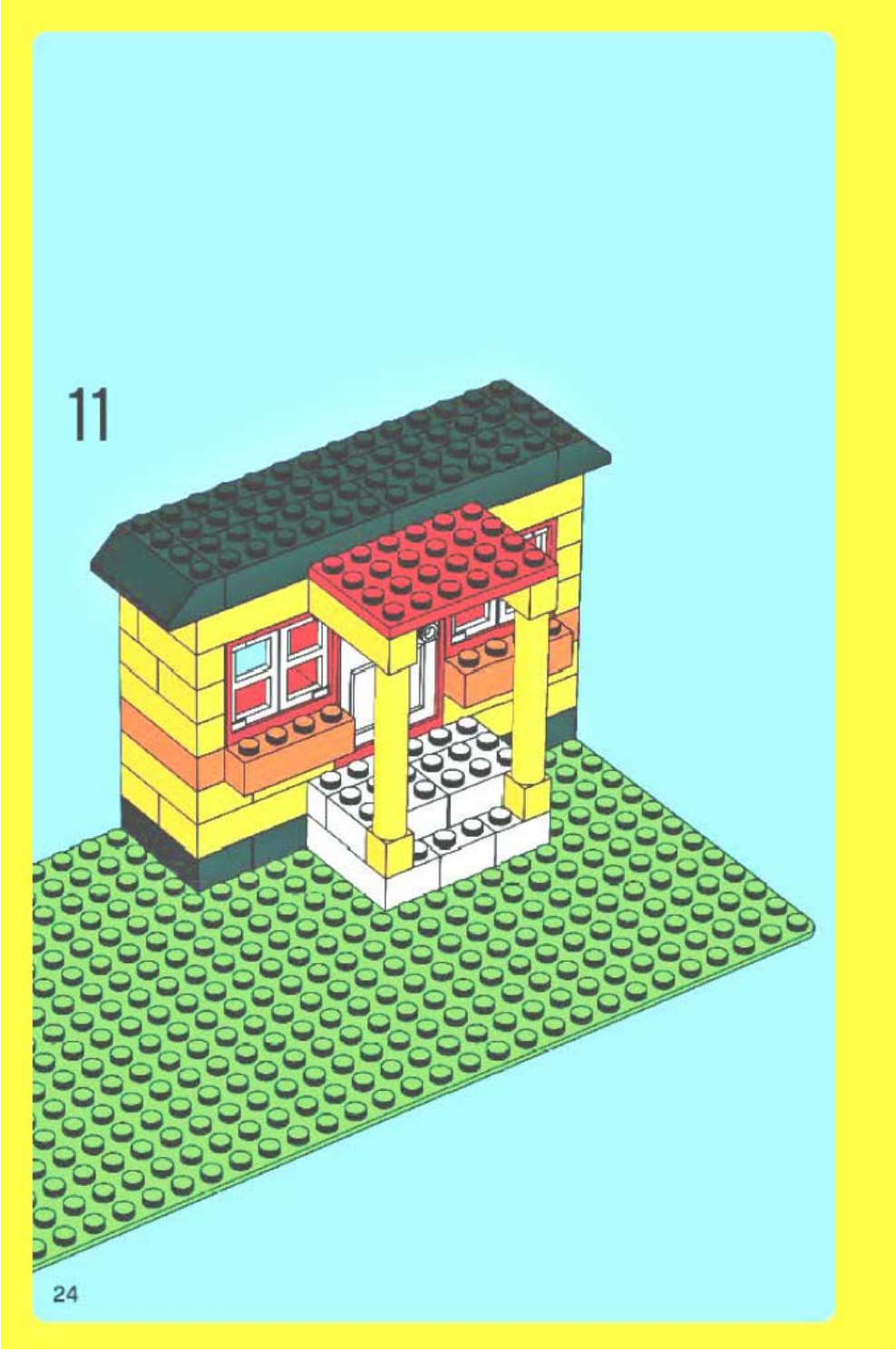

6
19

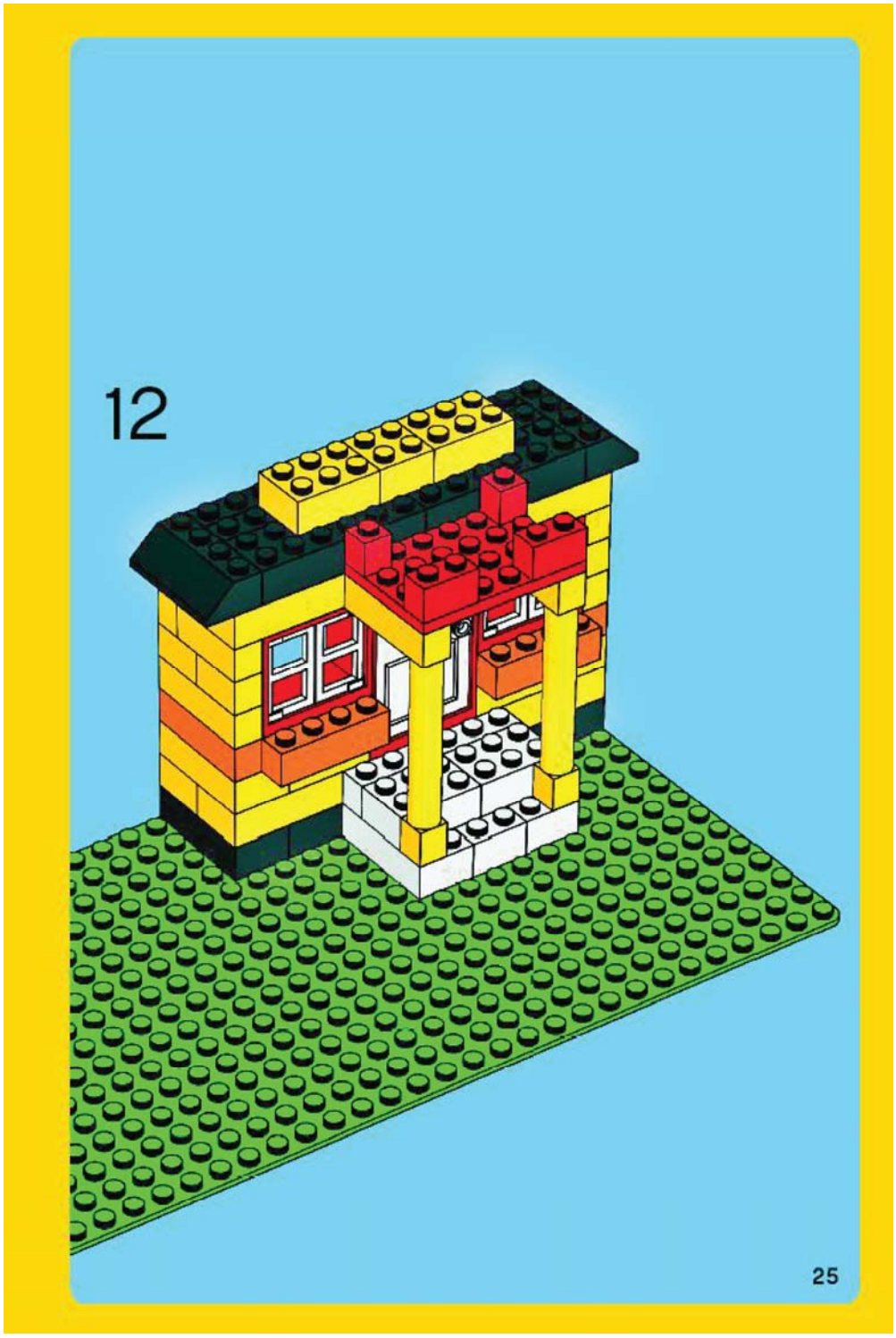

7
20

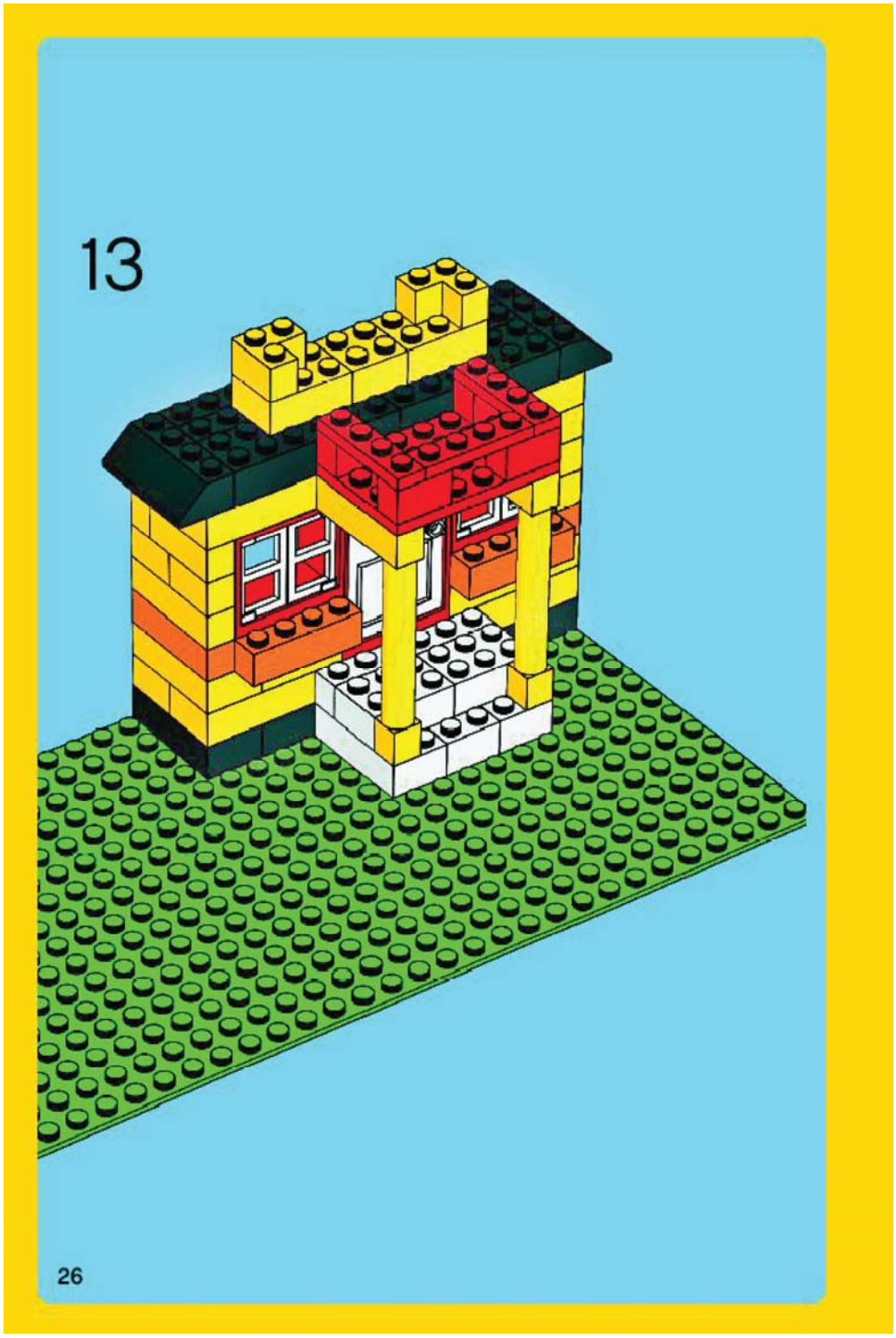

2x
8
21

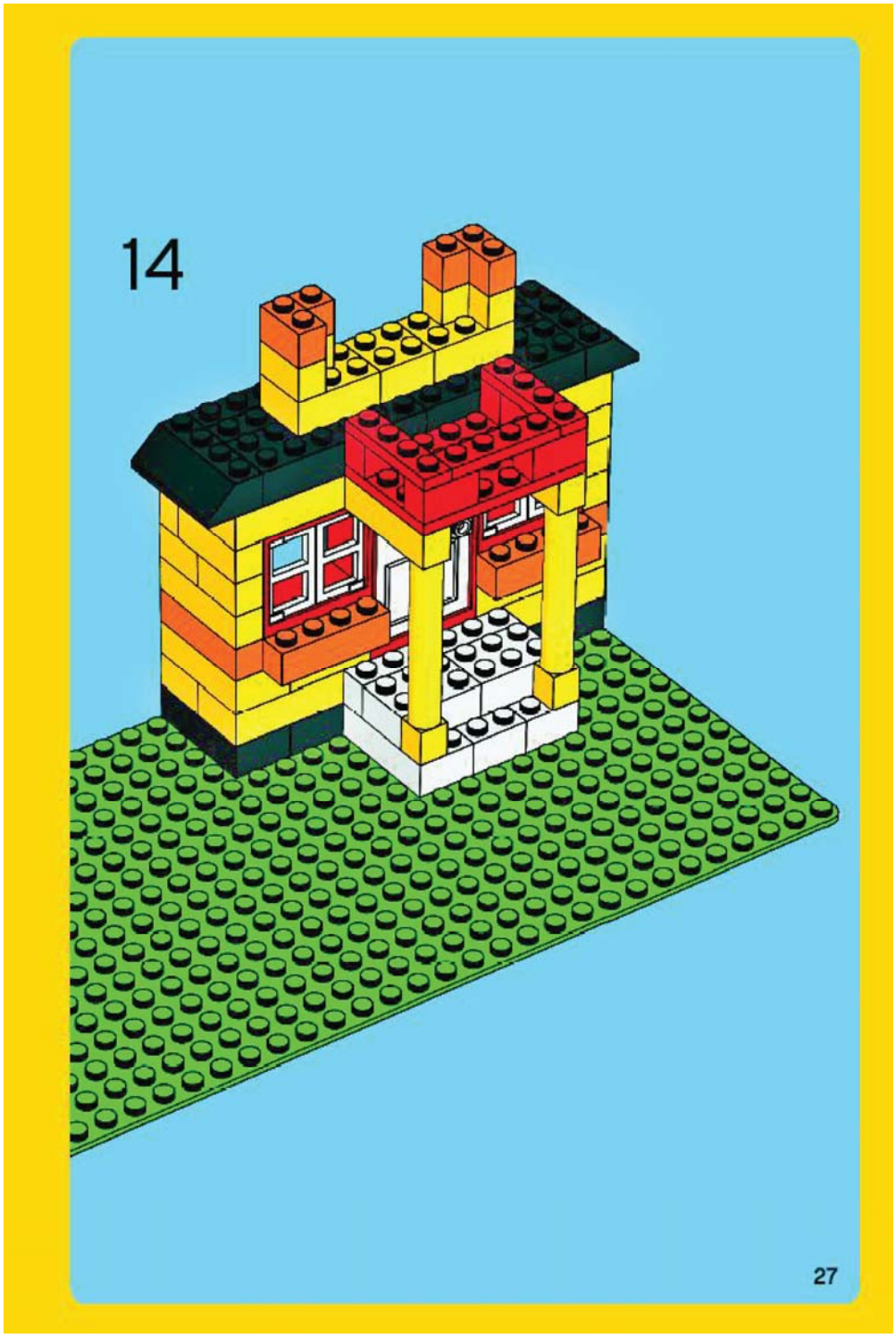

9
22

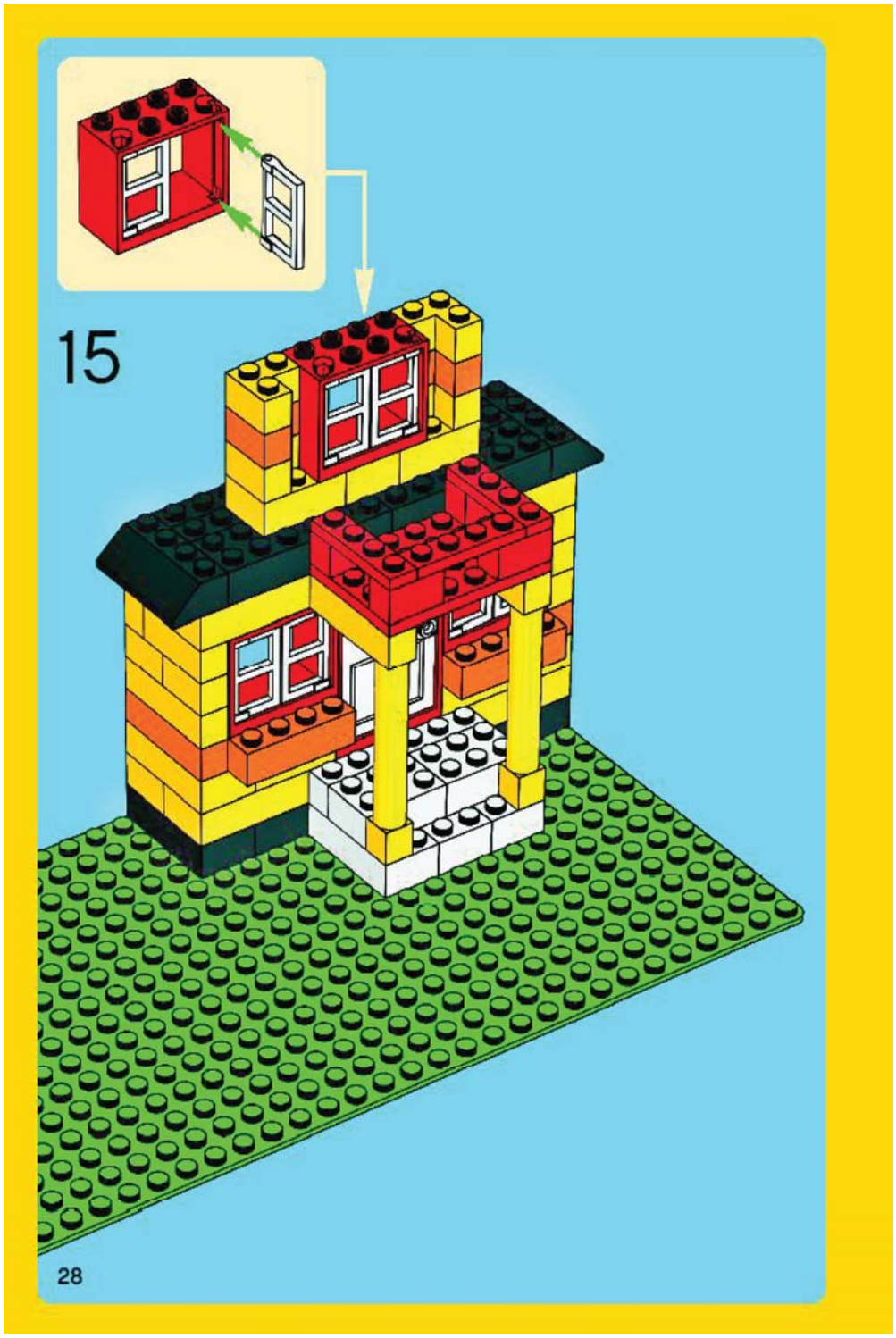

10
23

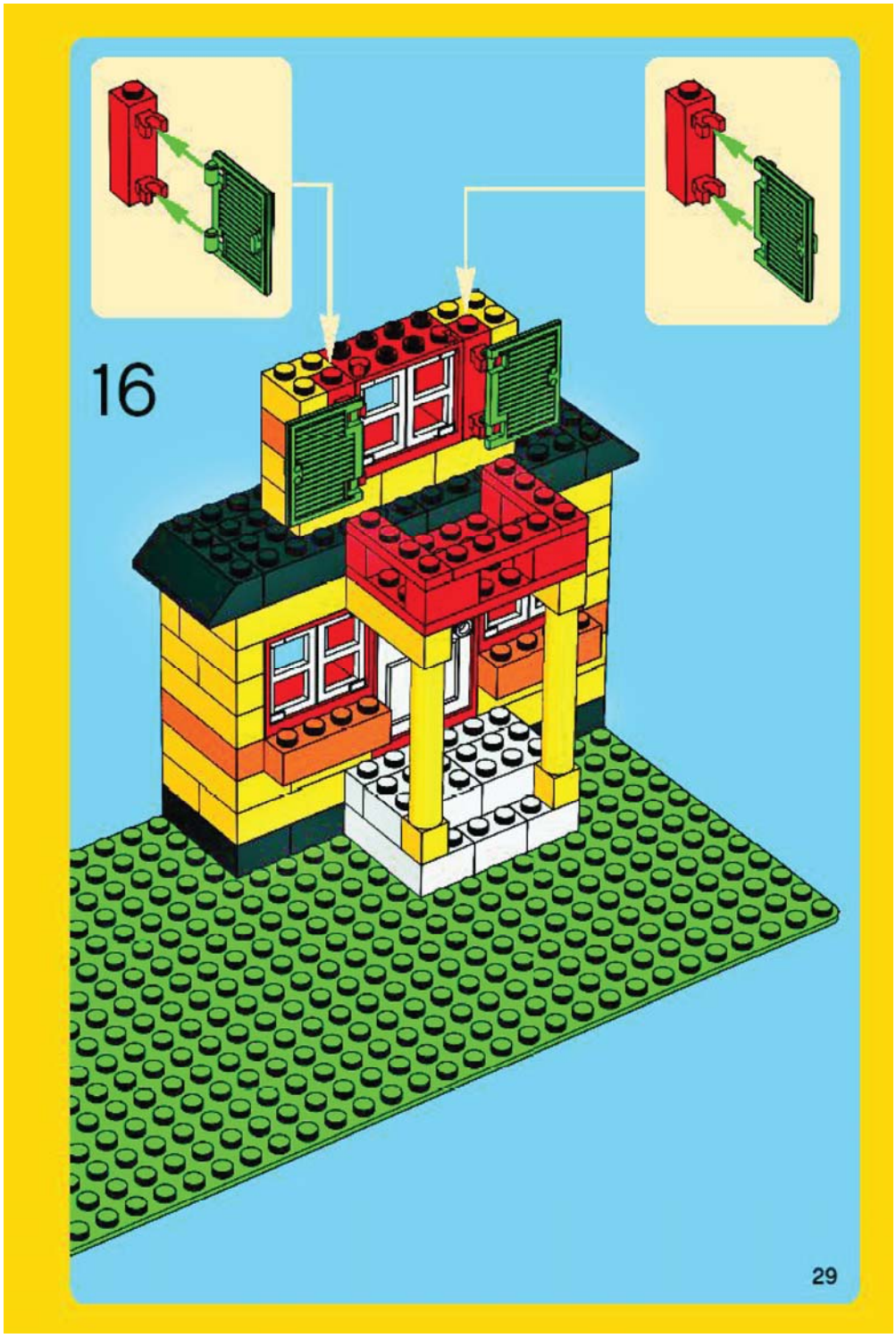

11
24

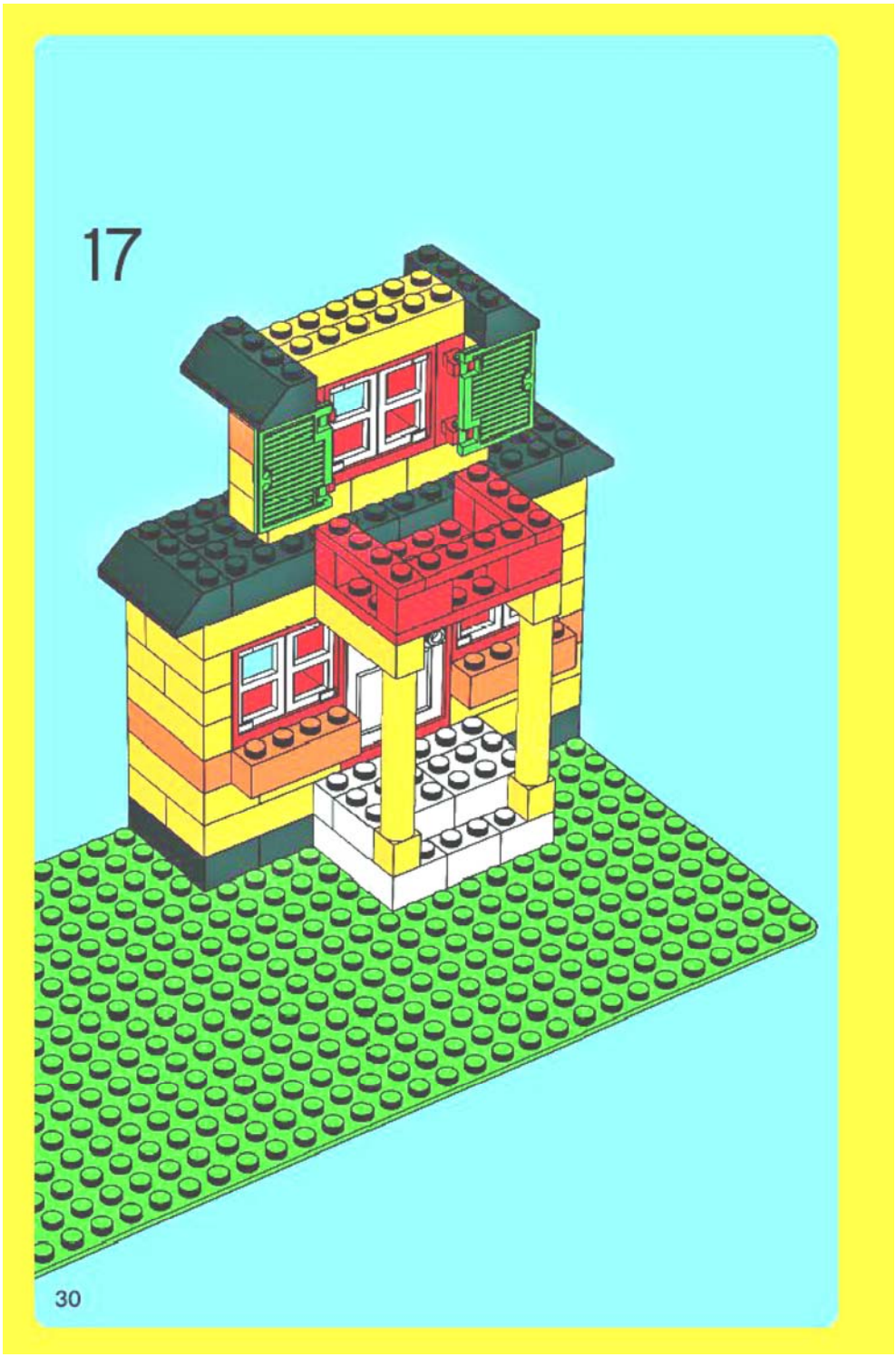

12
25


13
26


14
27

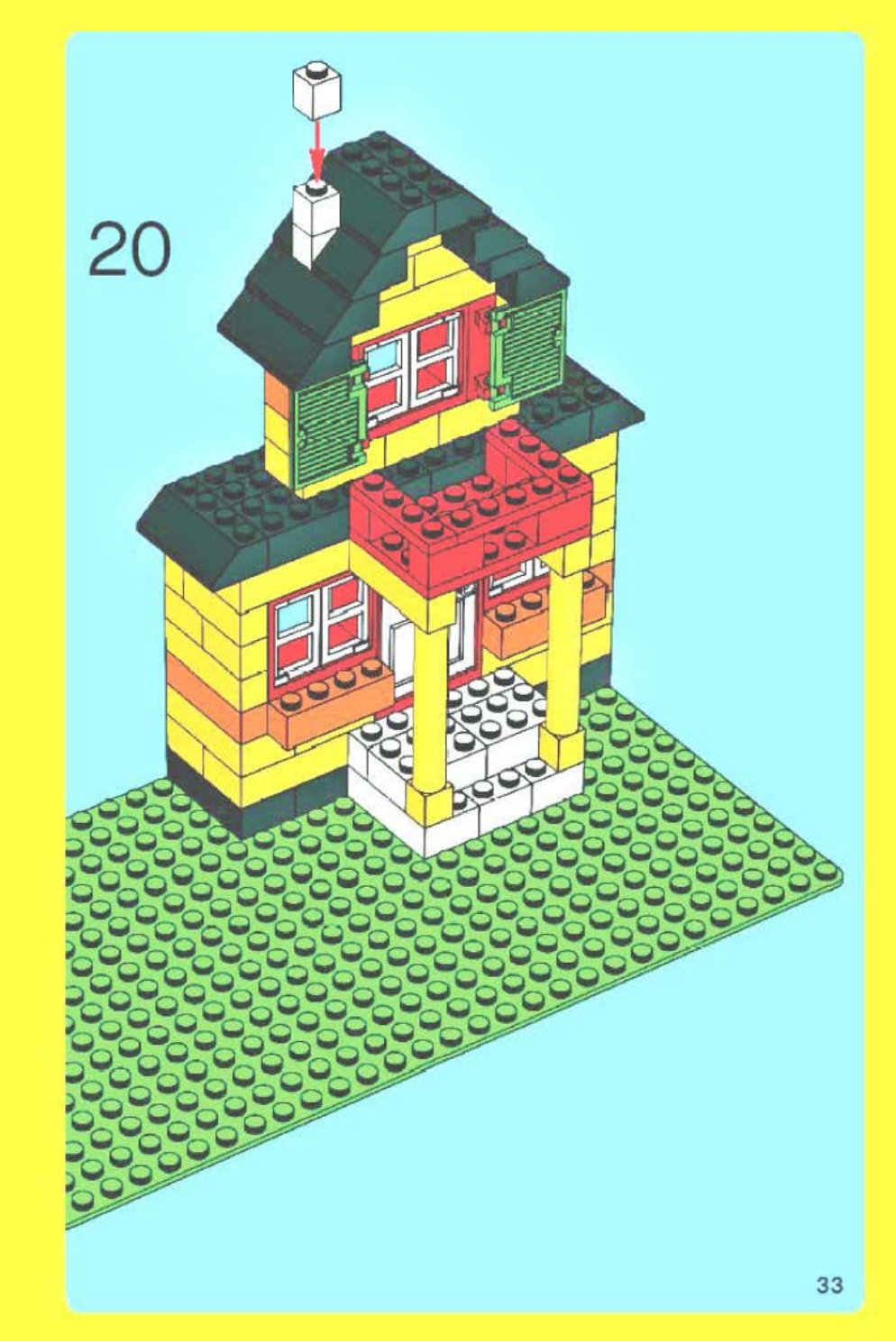

15
28


16
29

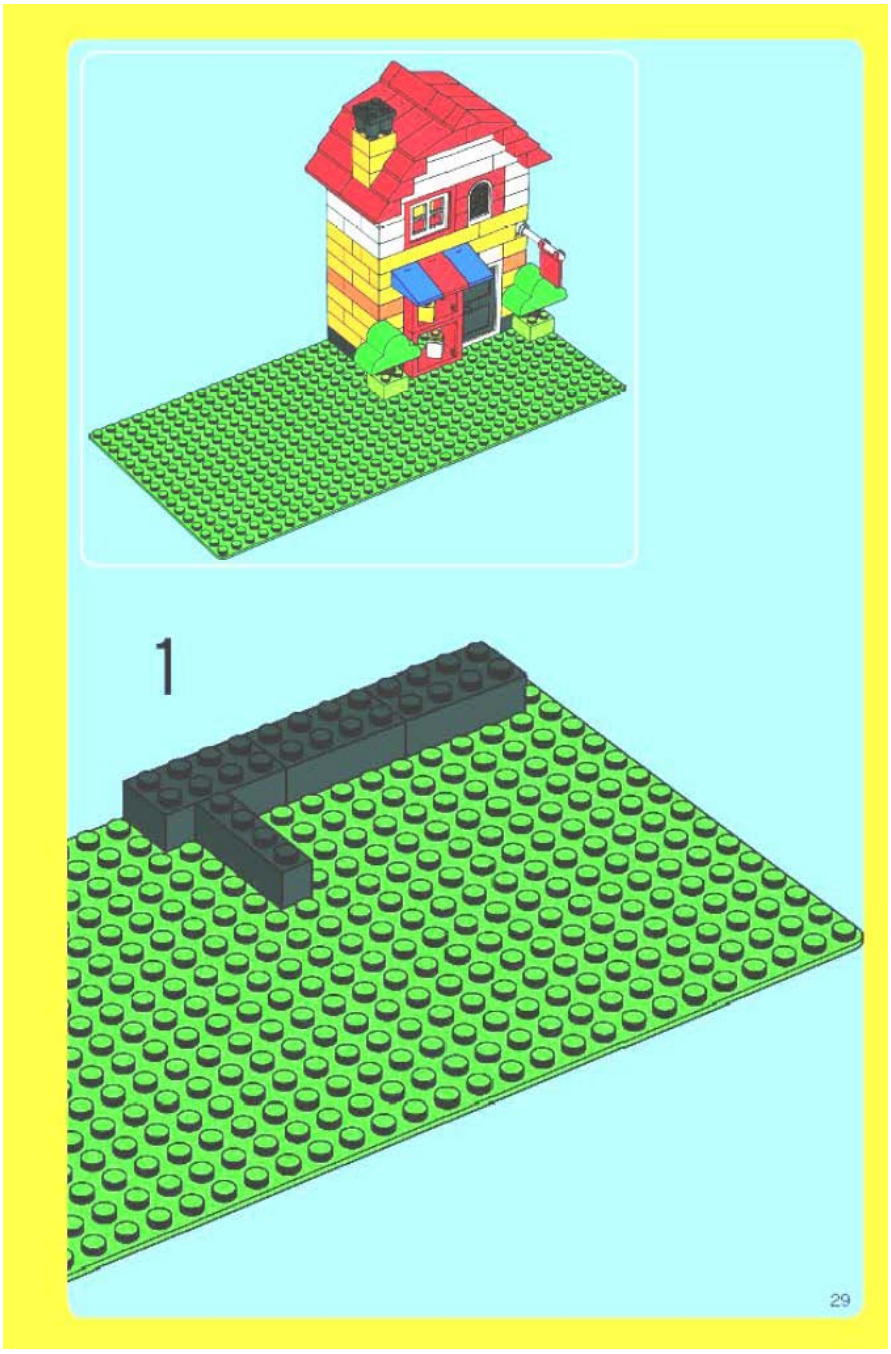

17
30

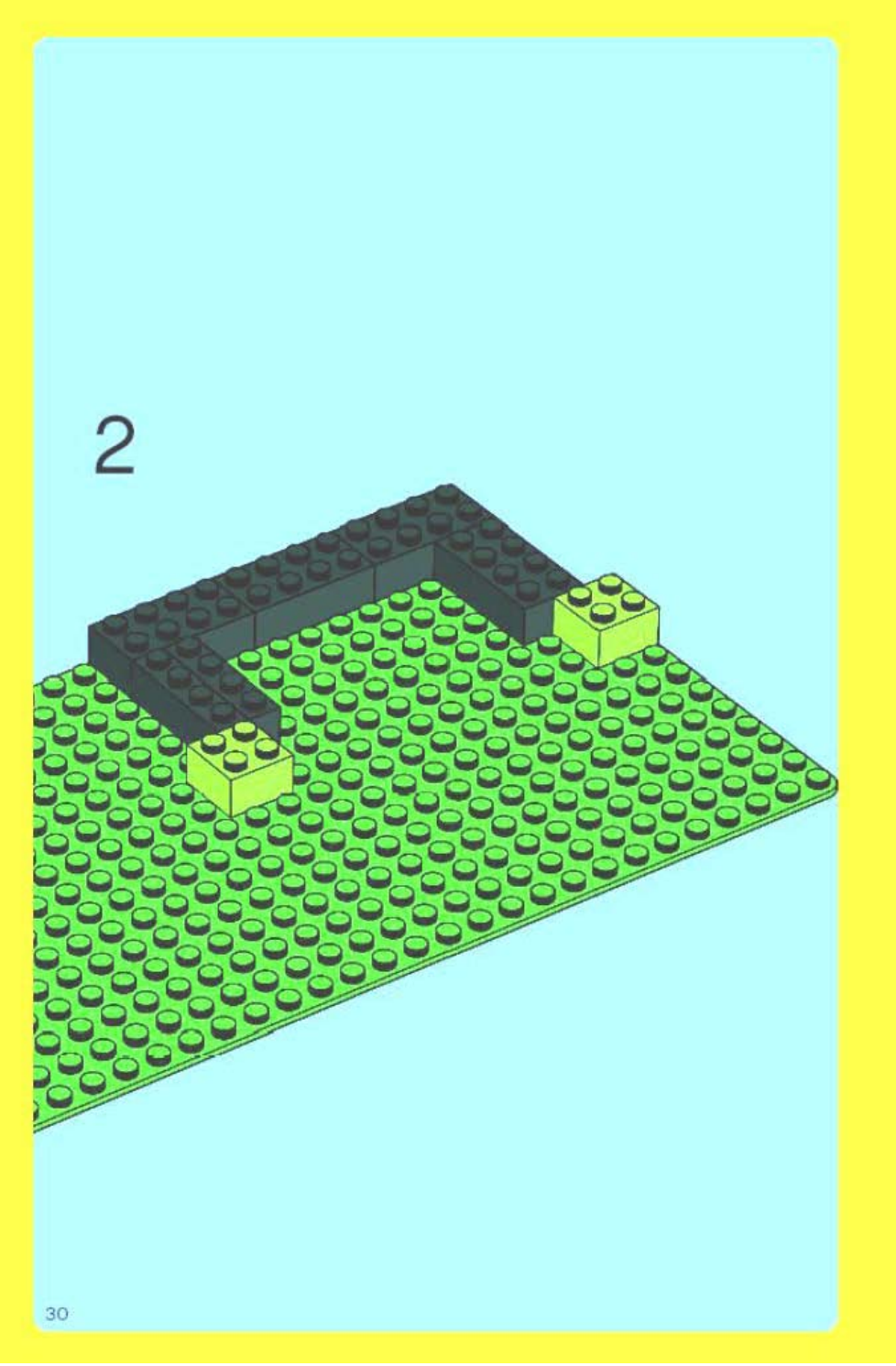
18
31

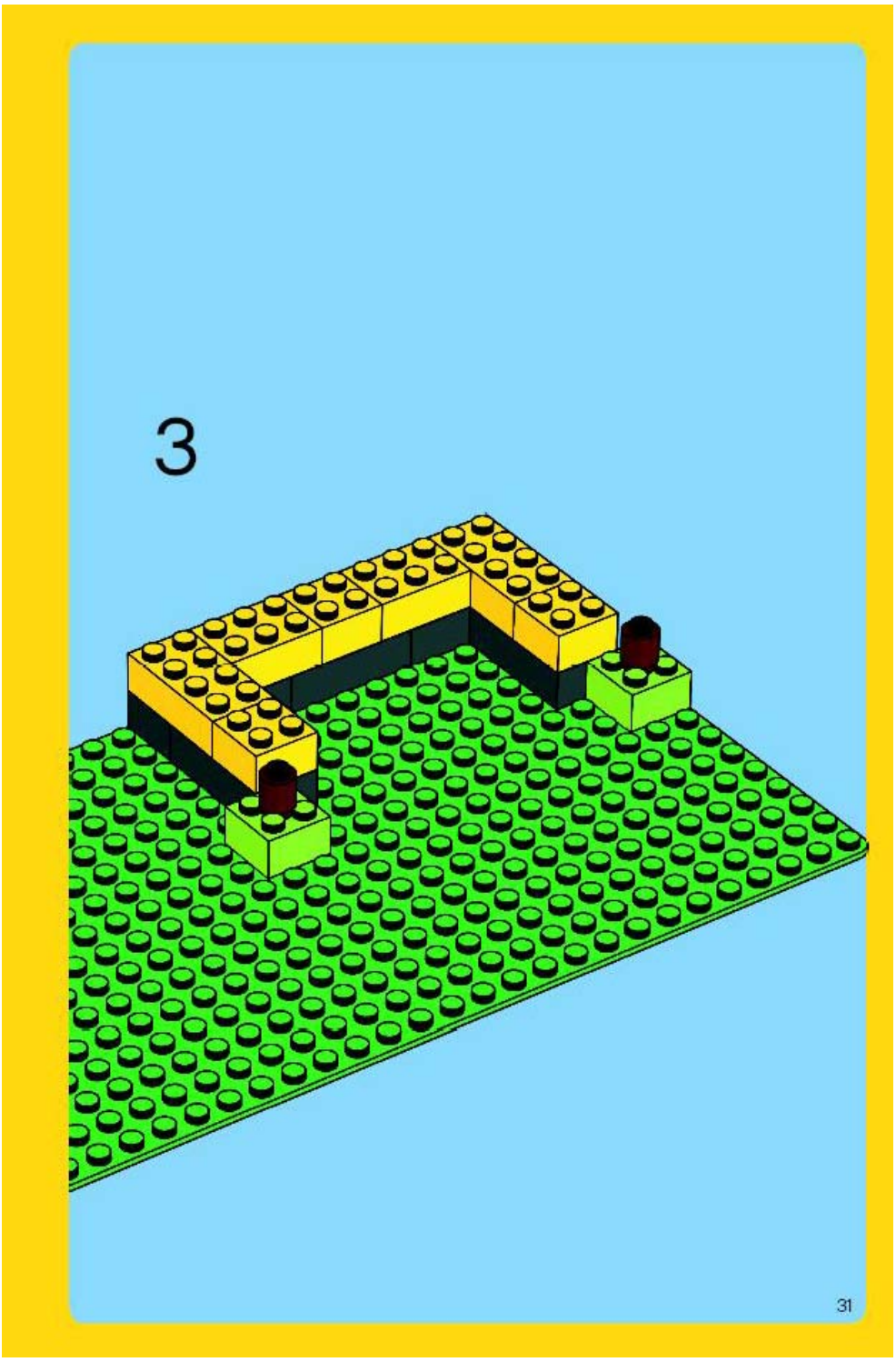
19
32

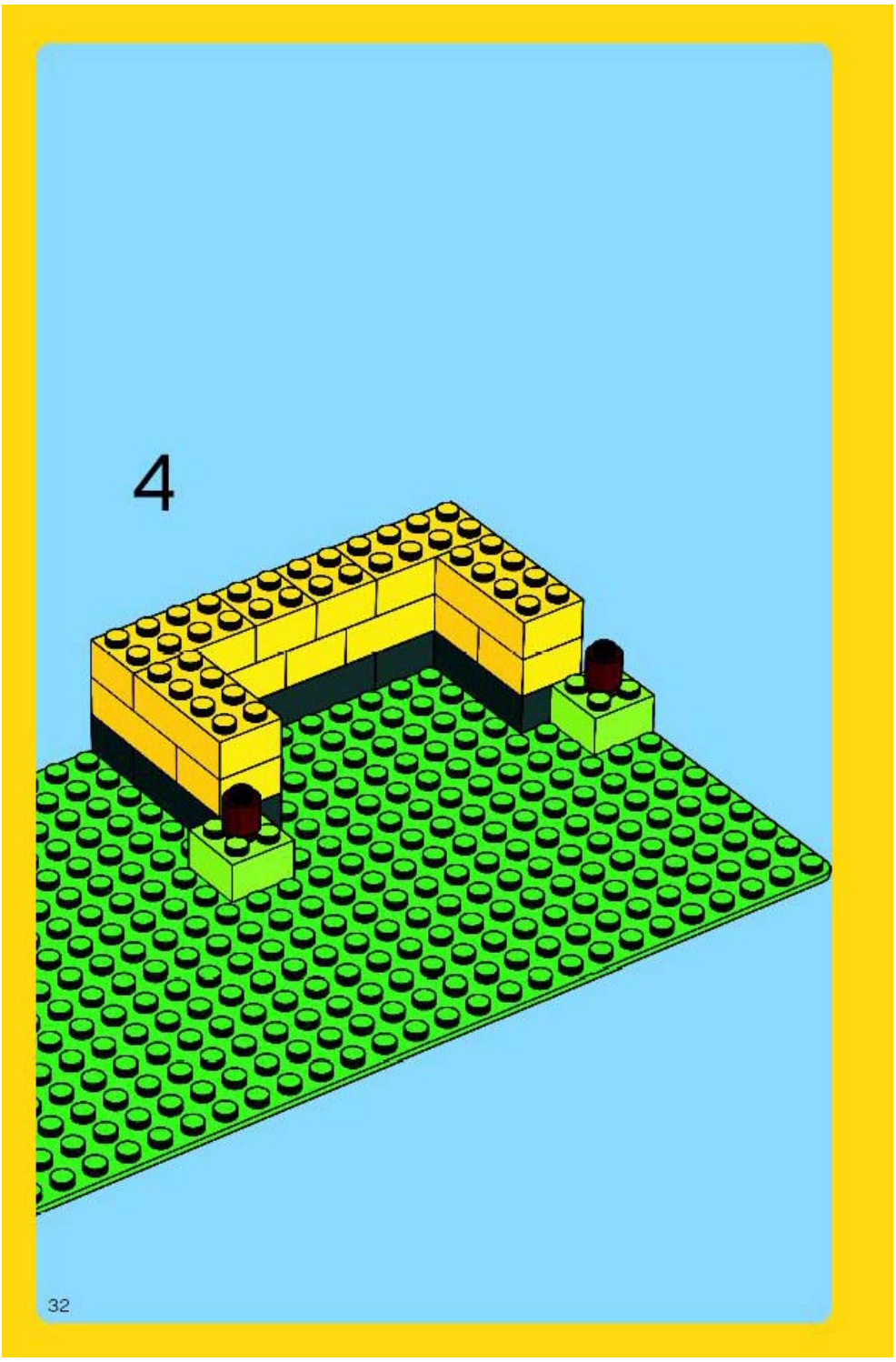

20
33

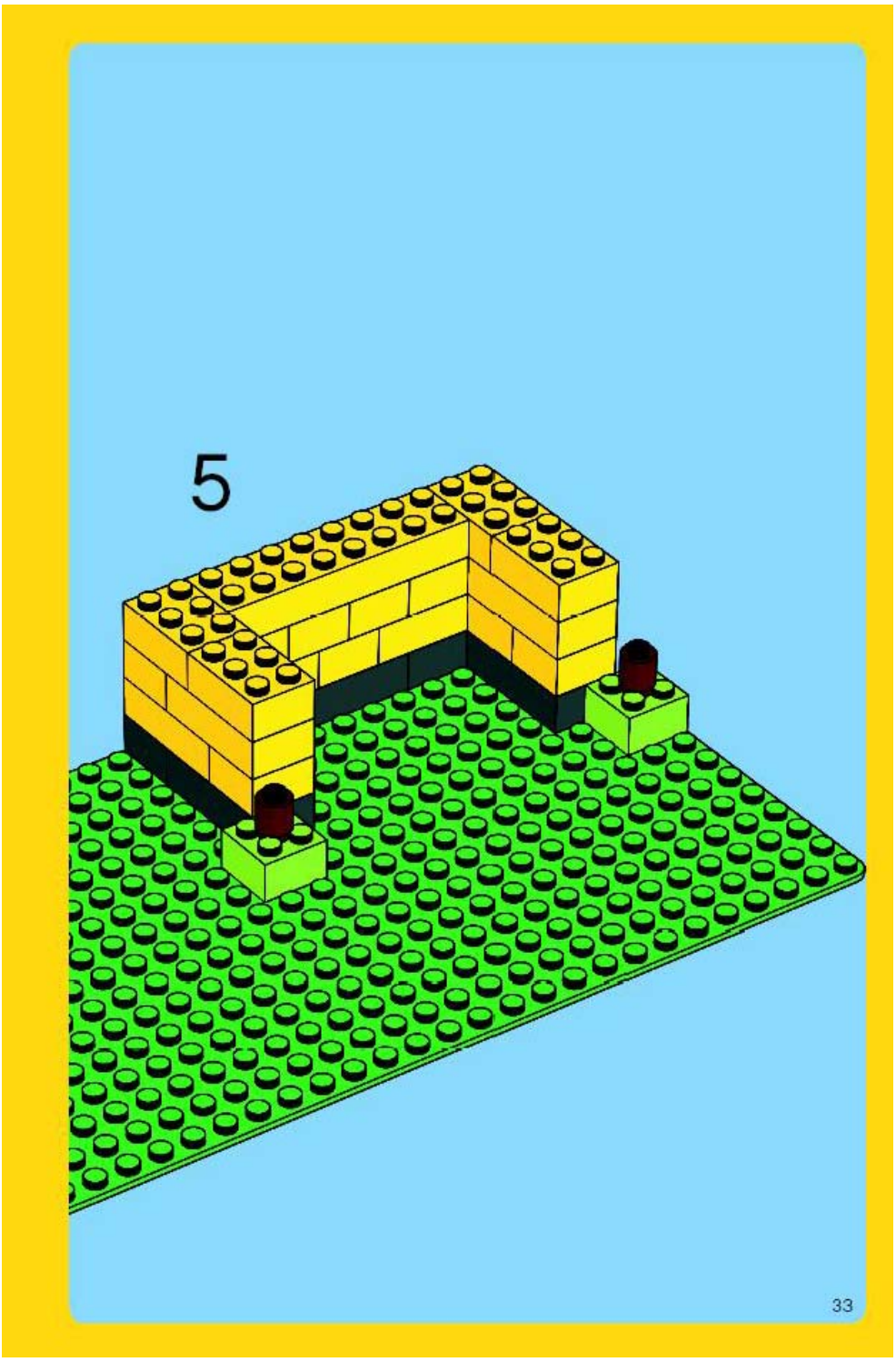
Appendix I Assembly Instructions from Pattern B (Lego™ Kit 6167)

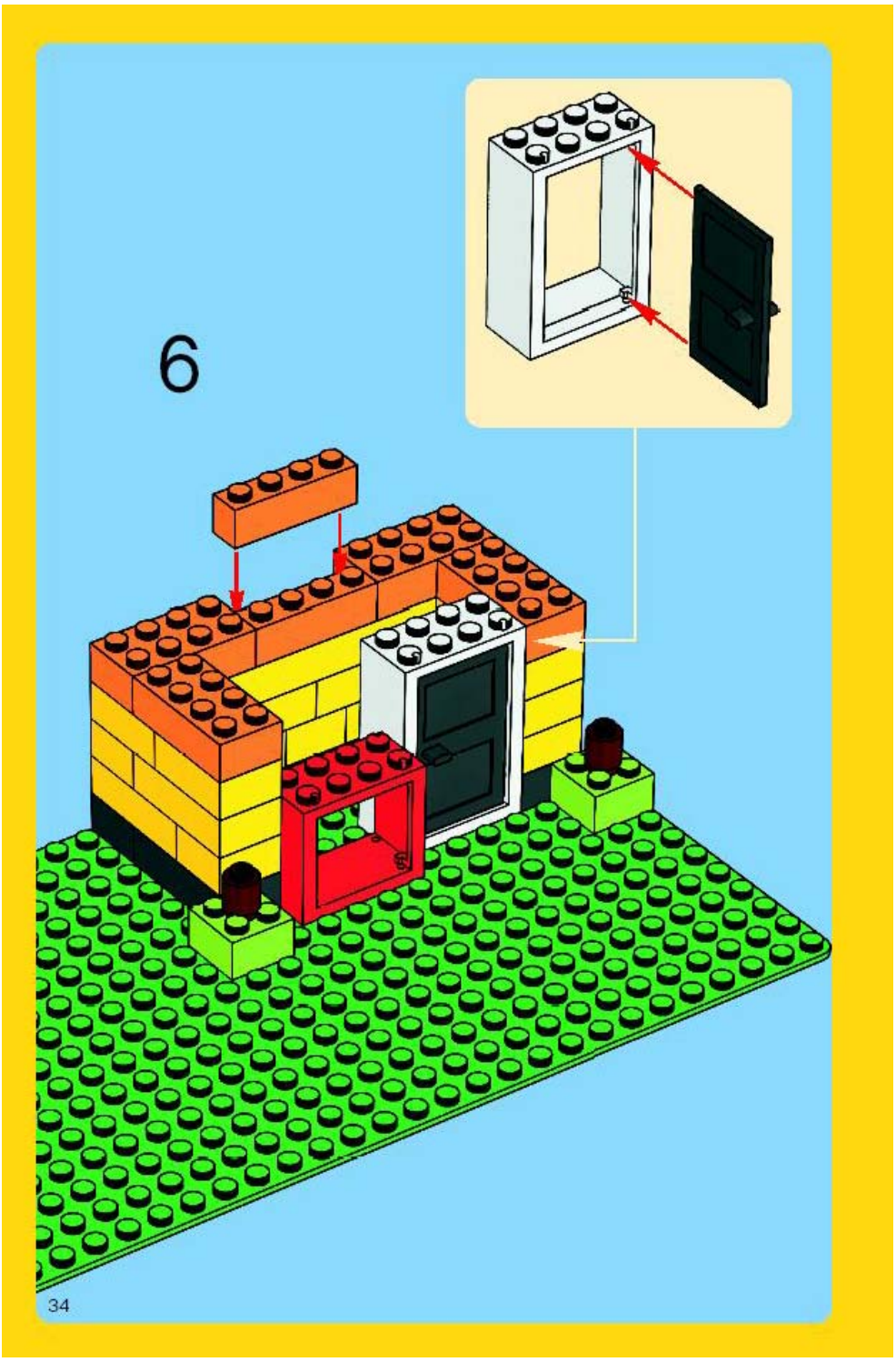


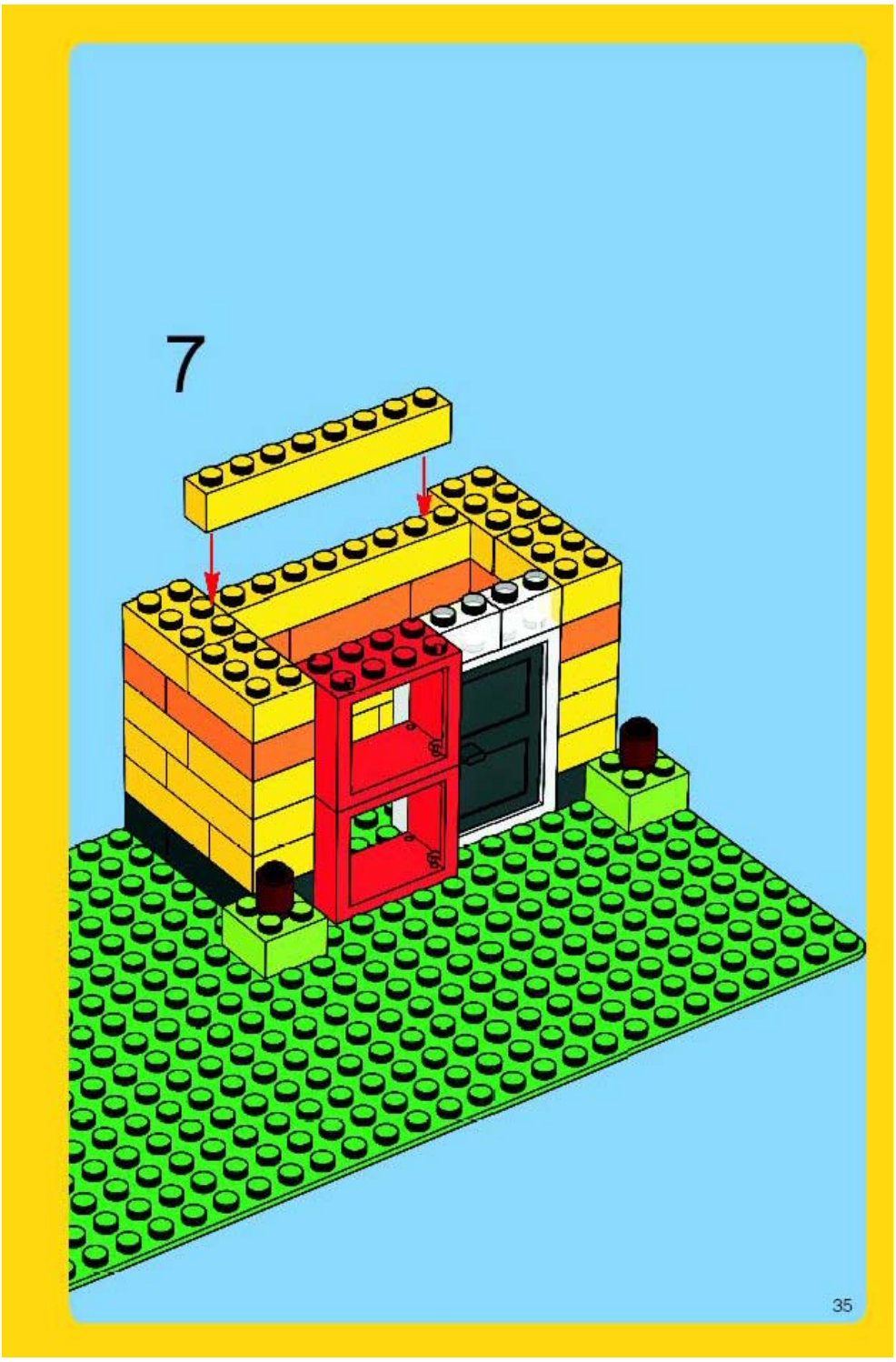

2
30

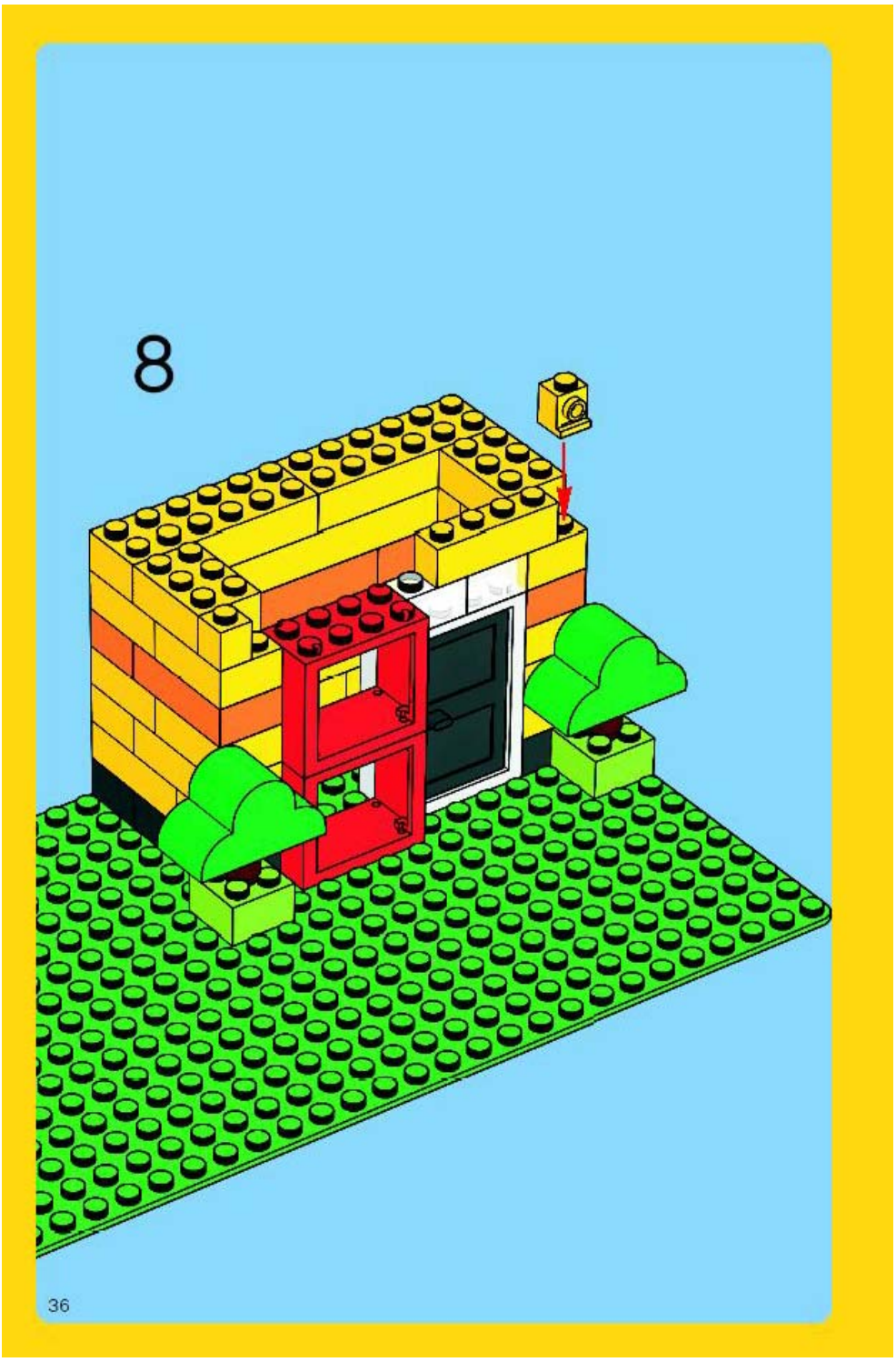

3
31

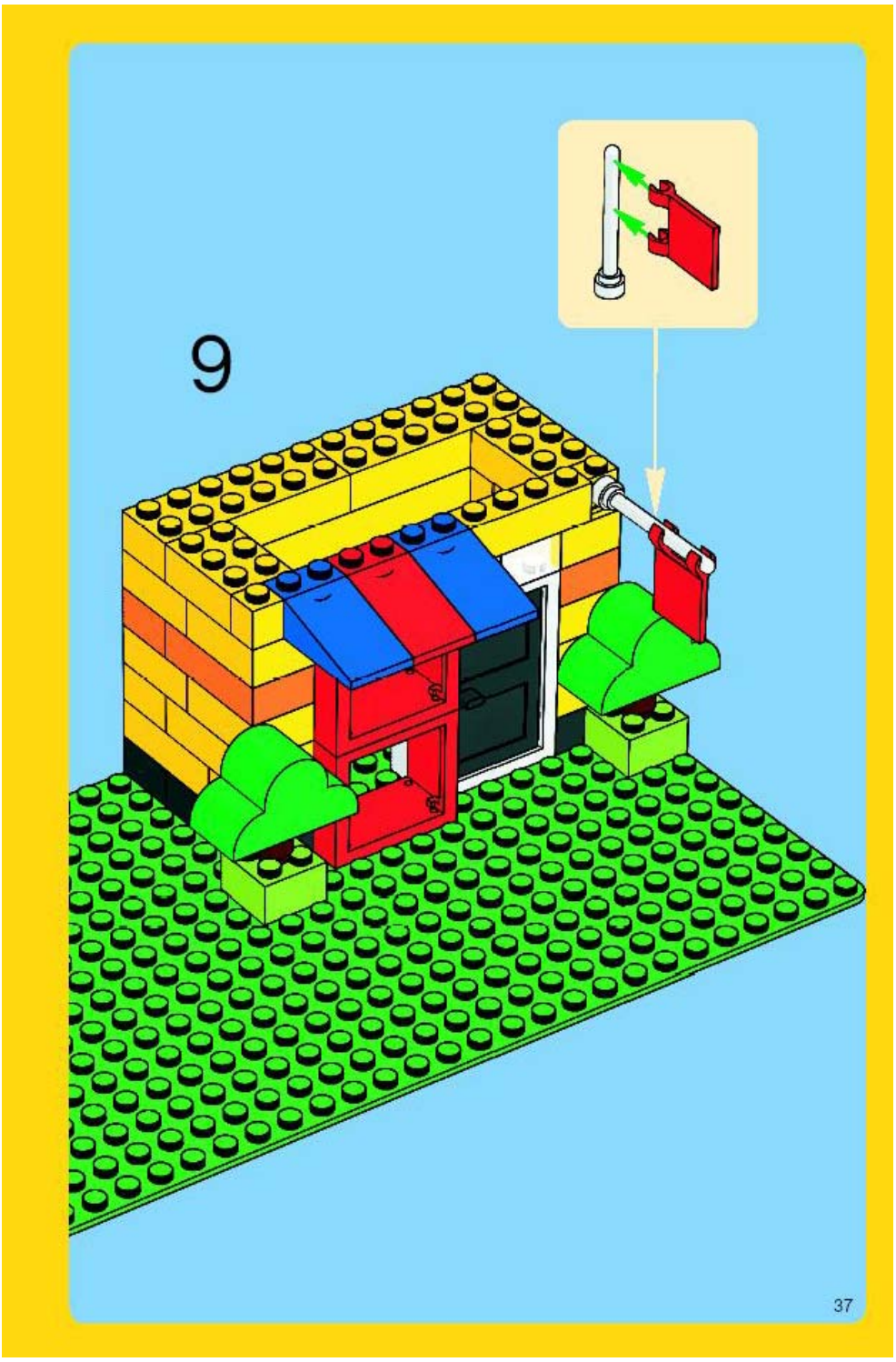

4
32

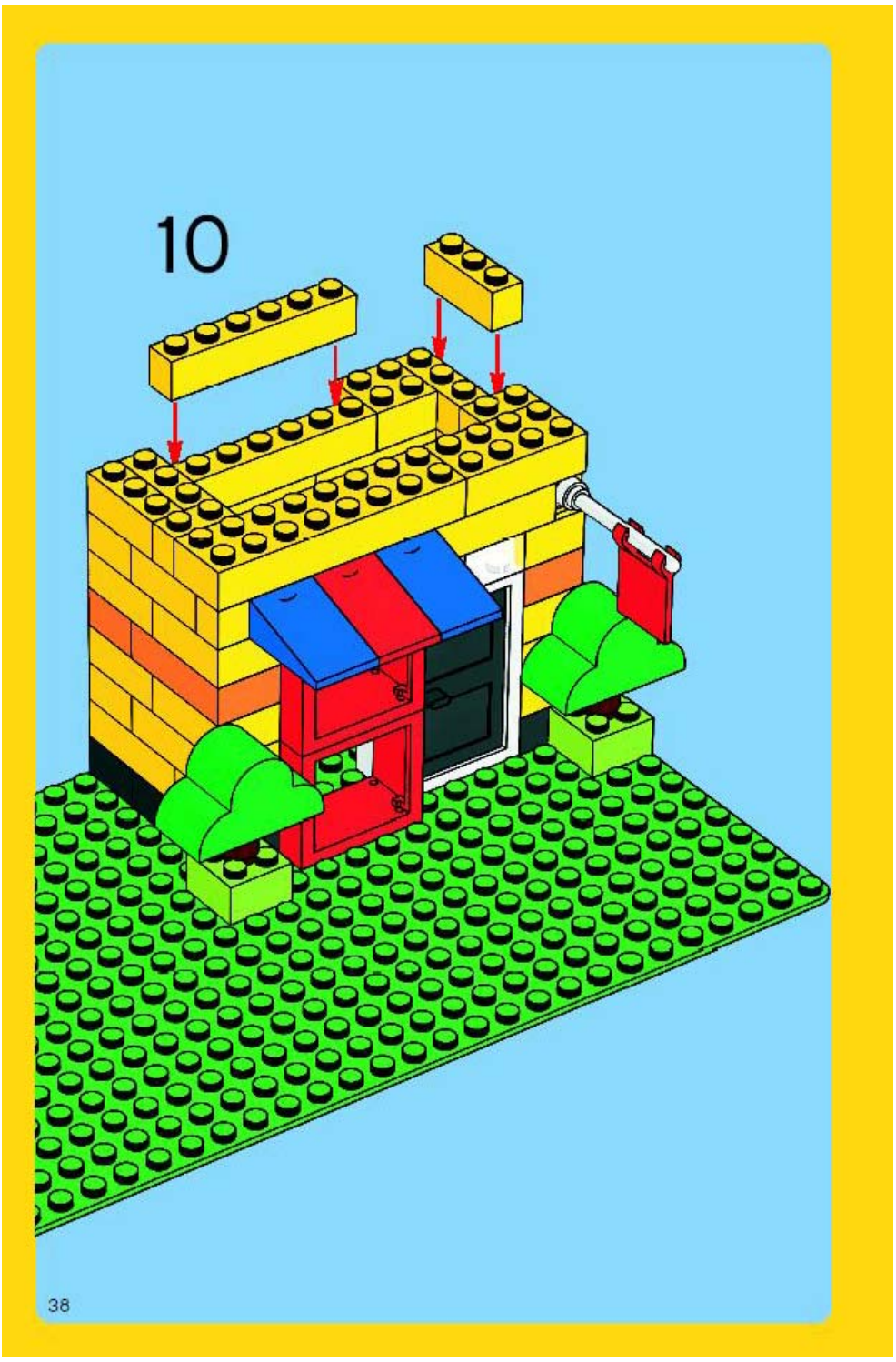

5
33

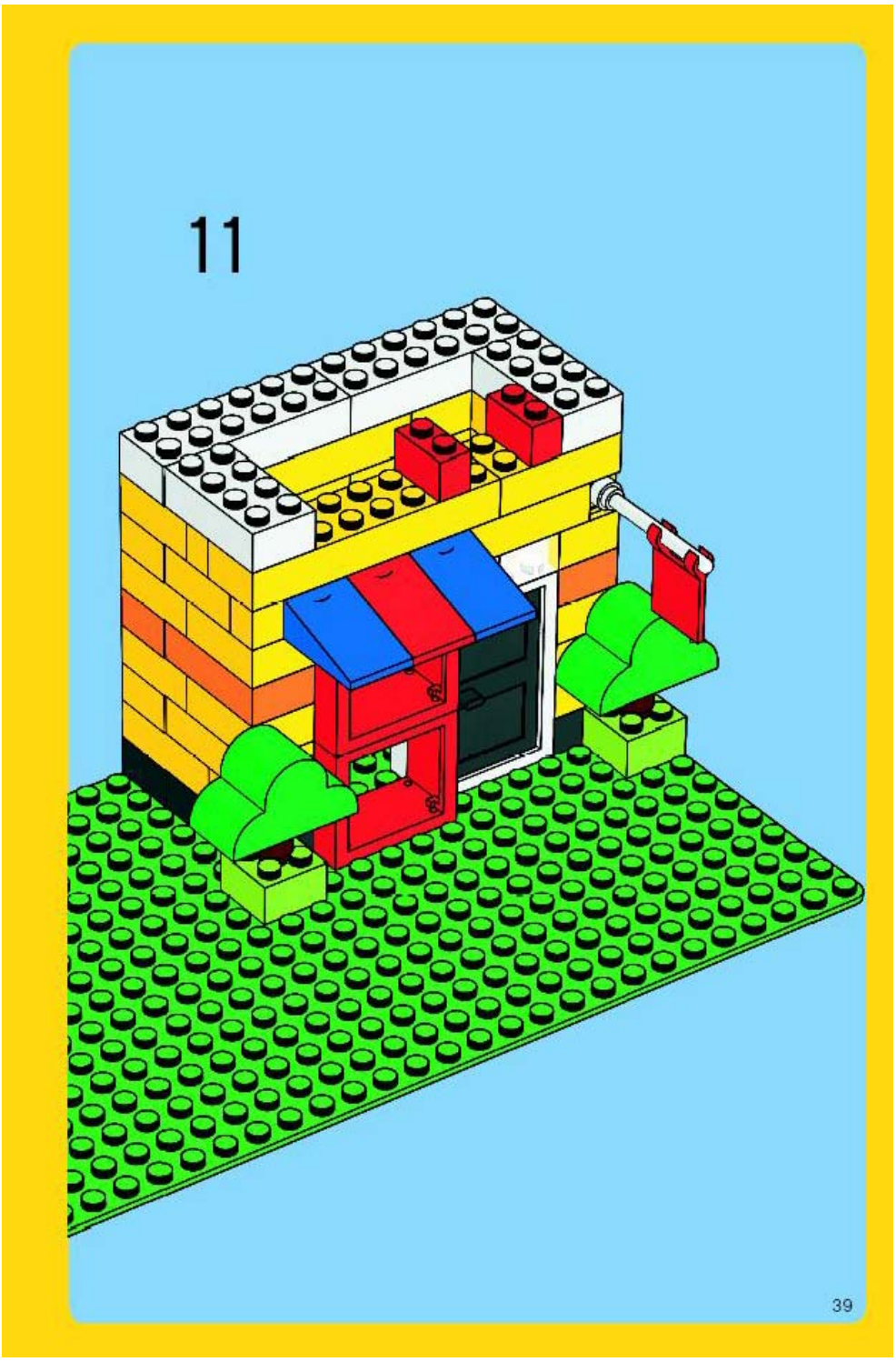

6
34

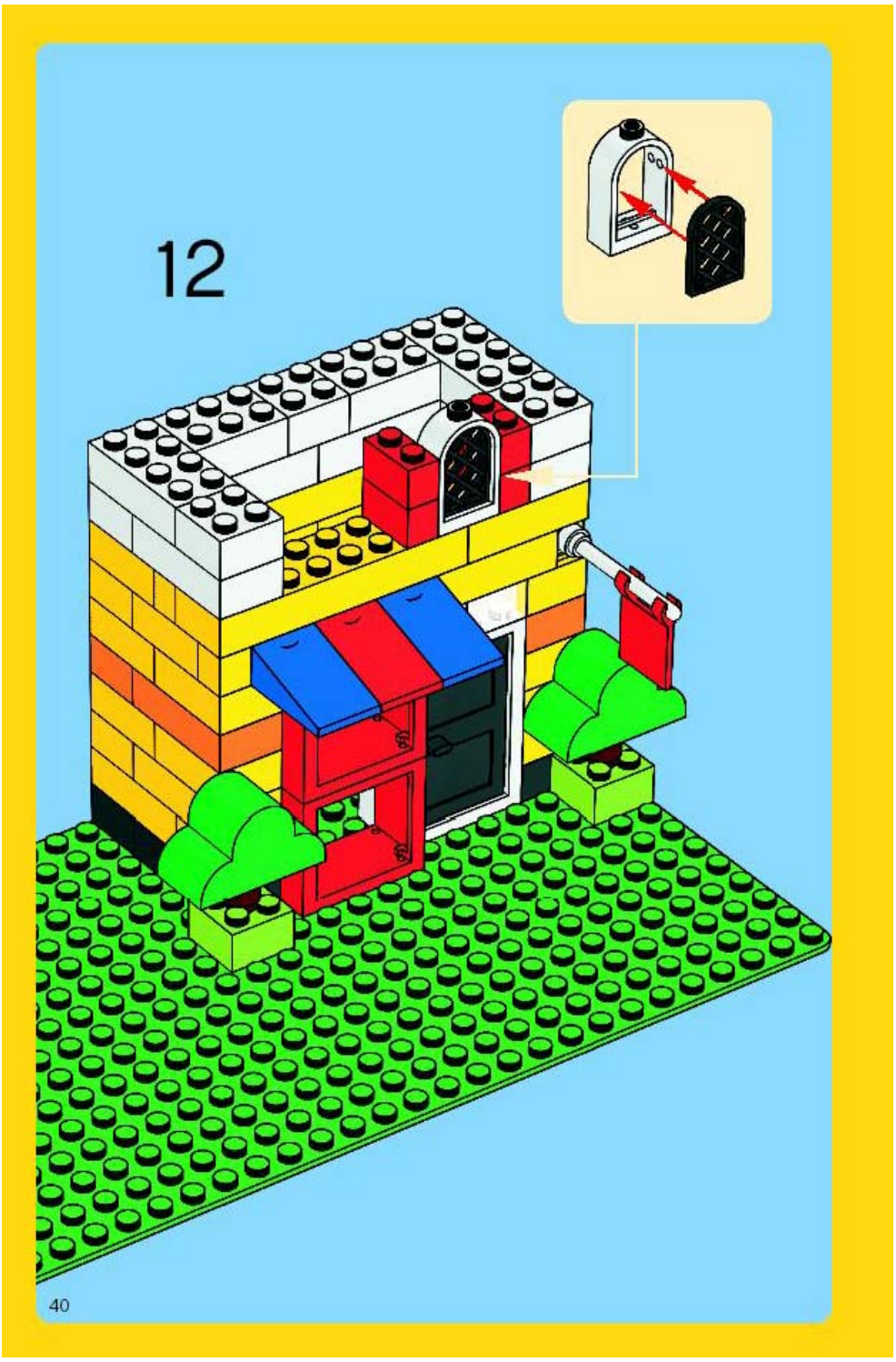

7
35

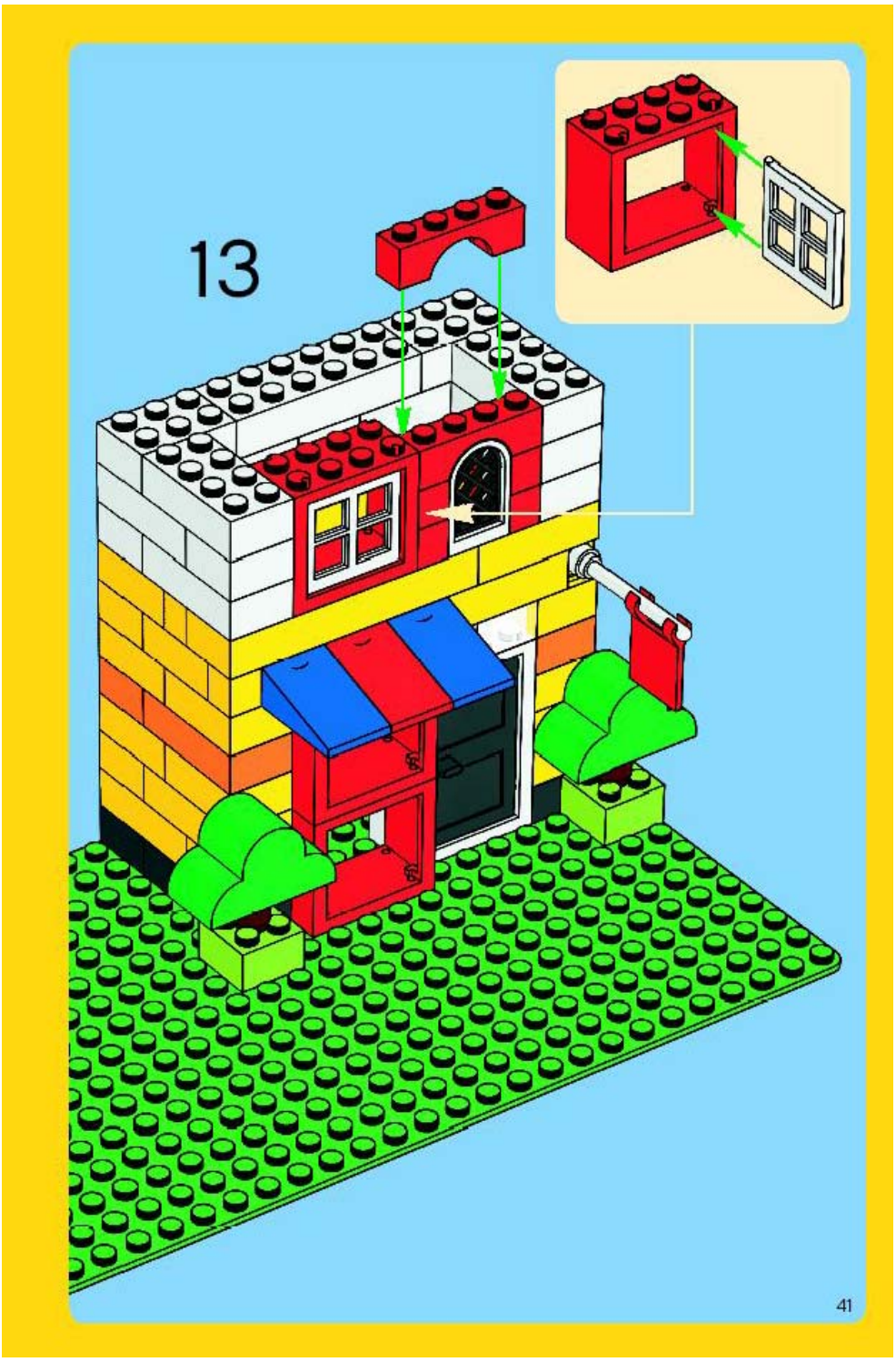

8
36

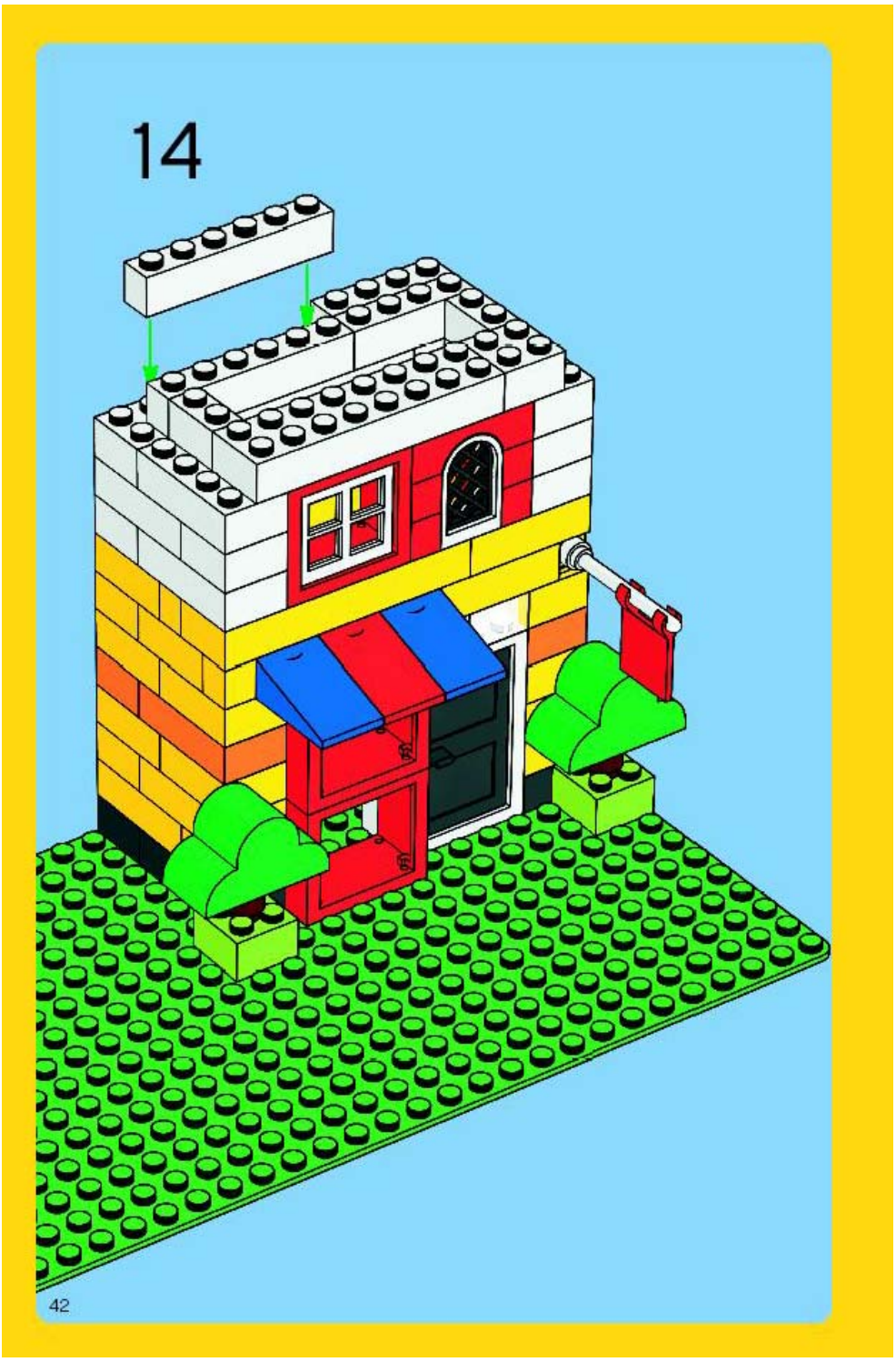

9
37

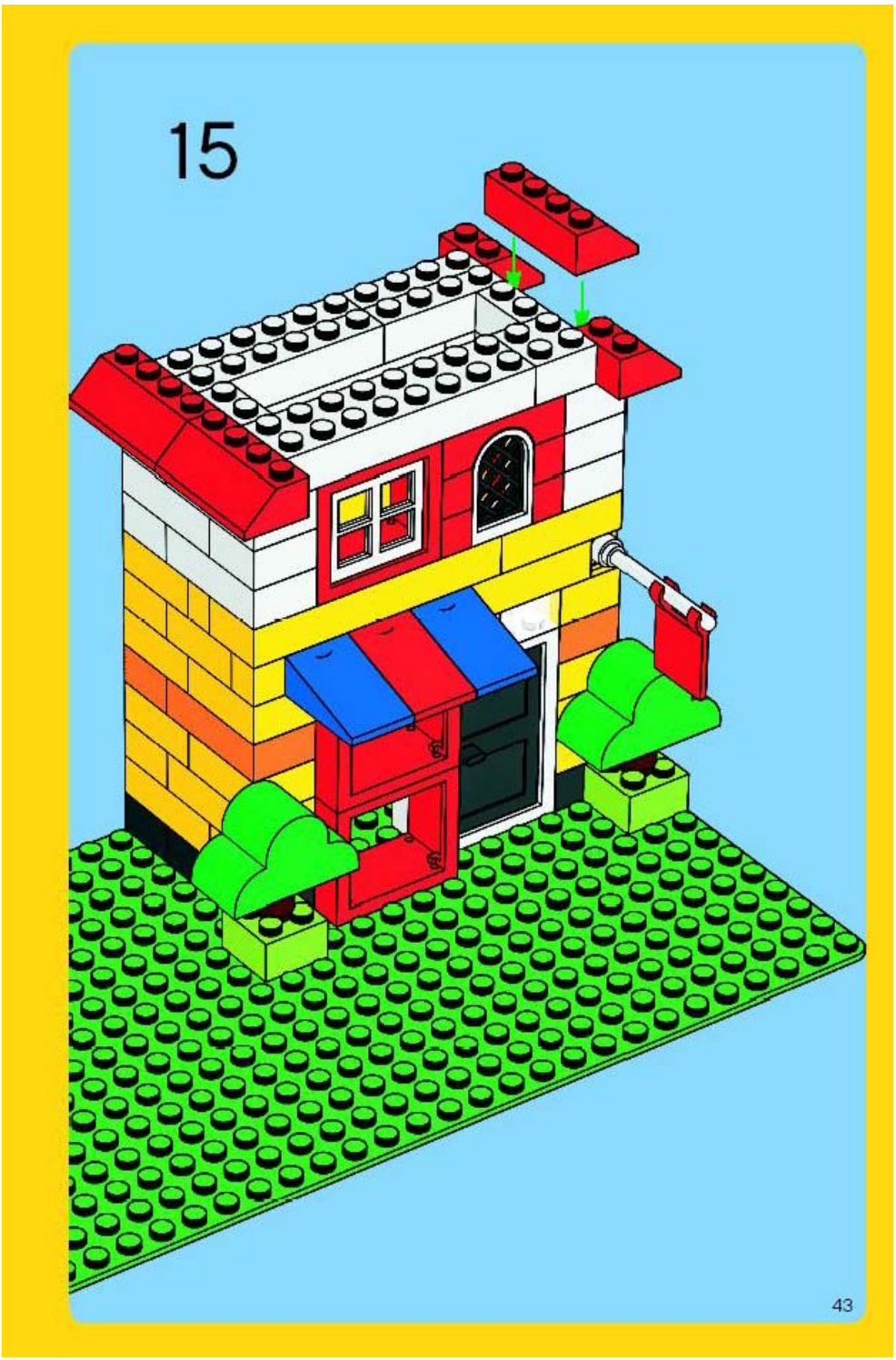

10
38

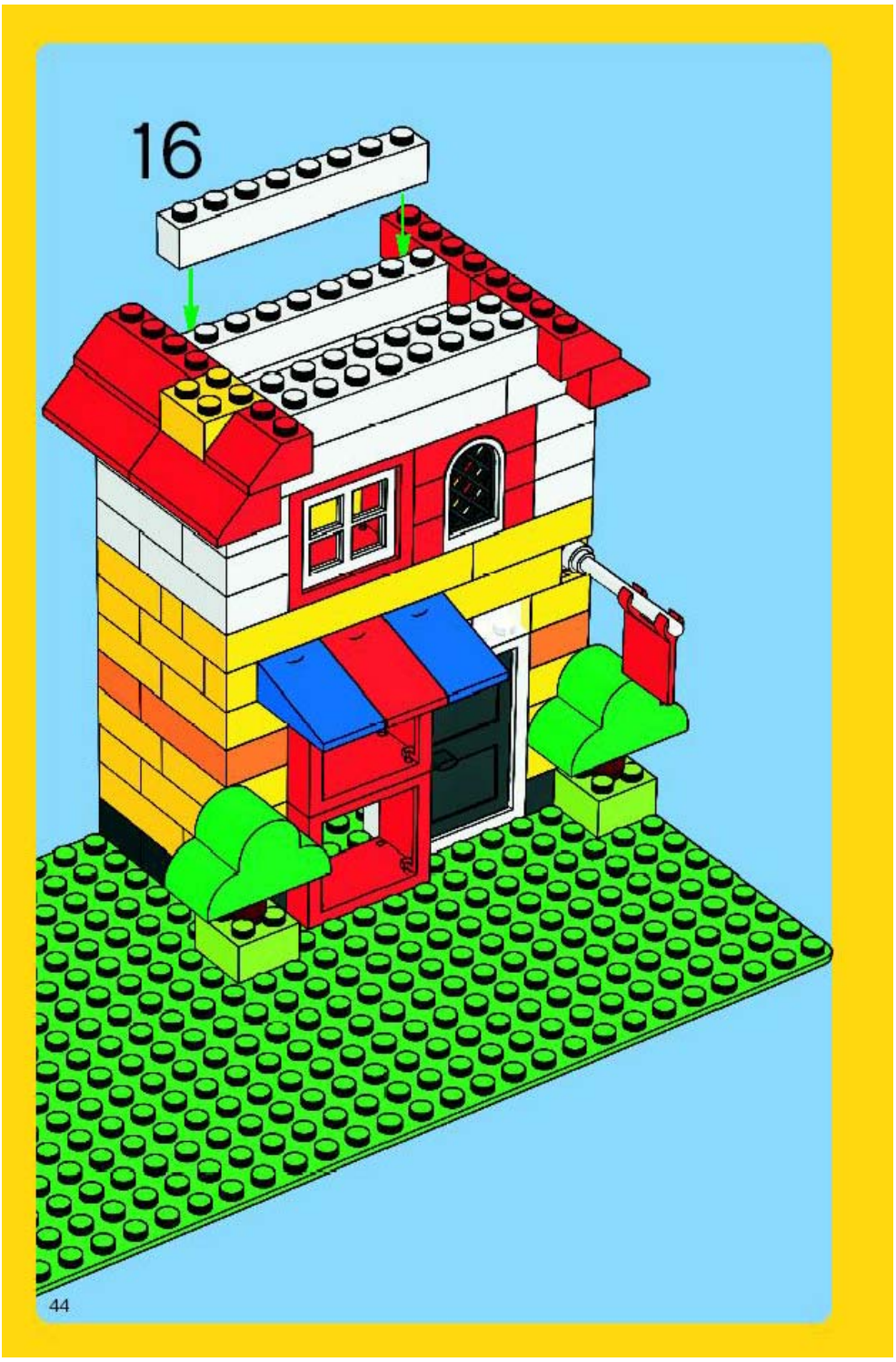

11
39

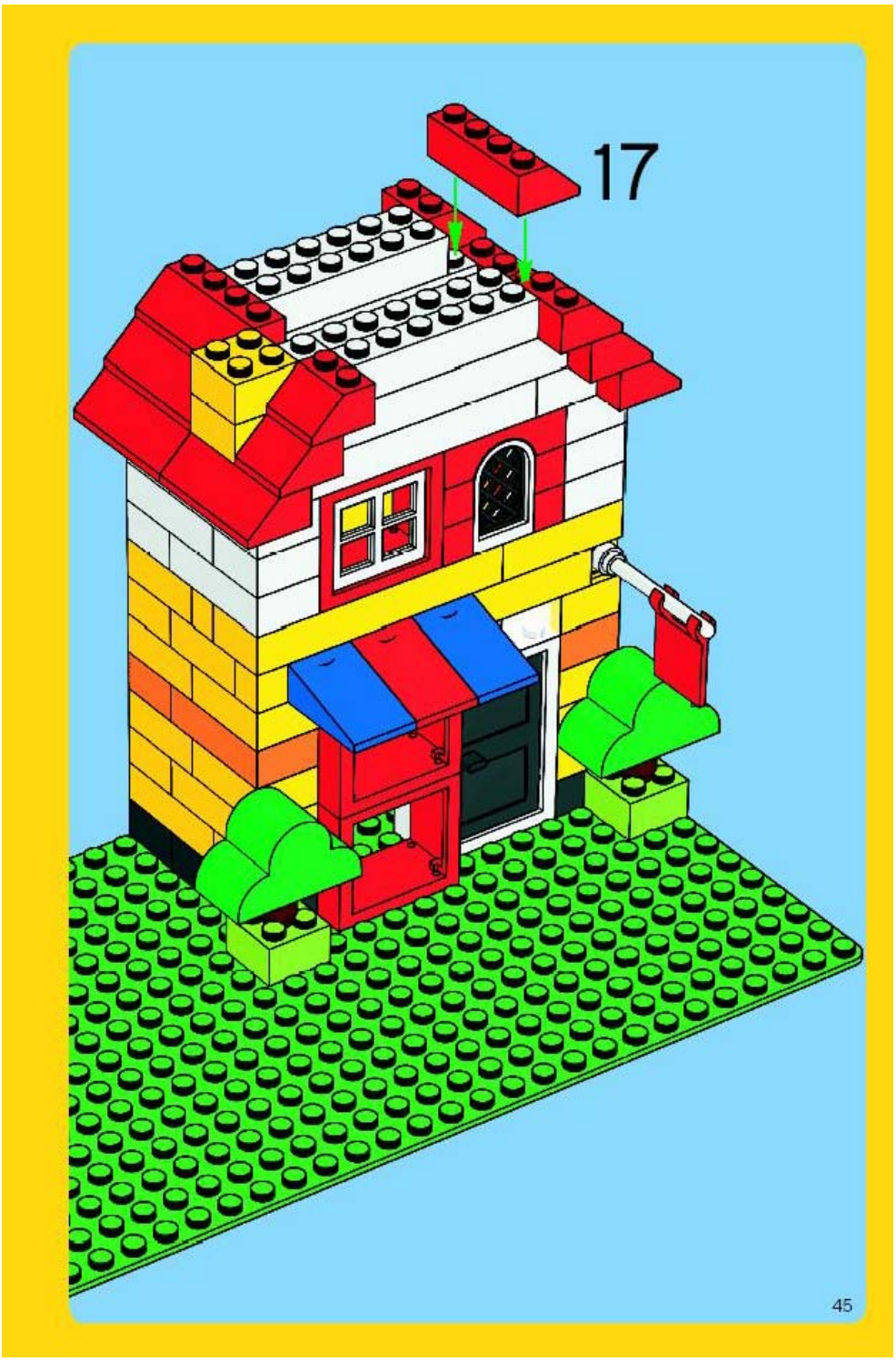

12
40


13
41

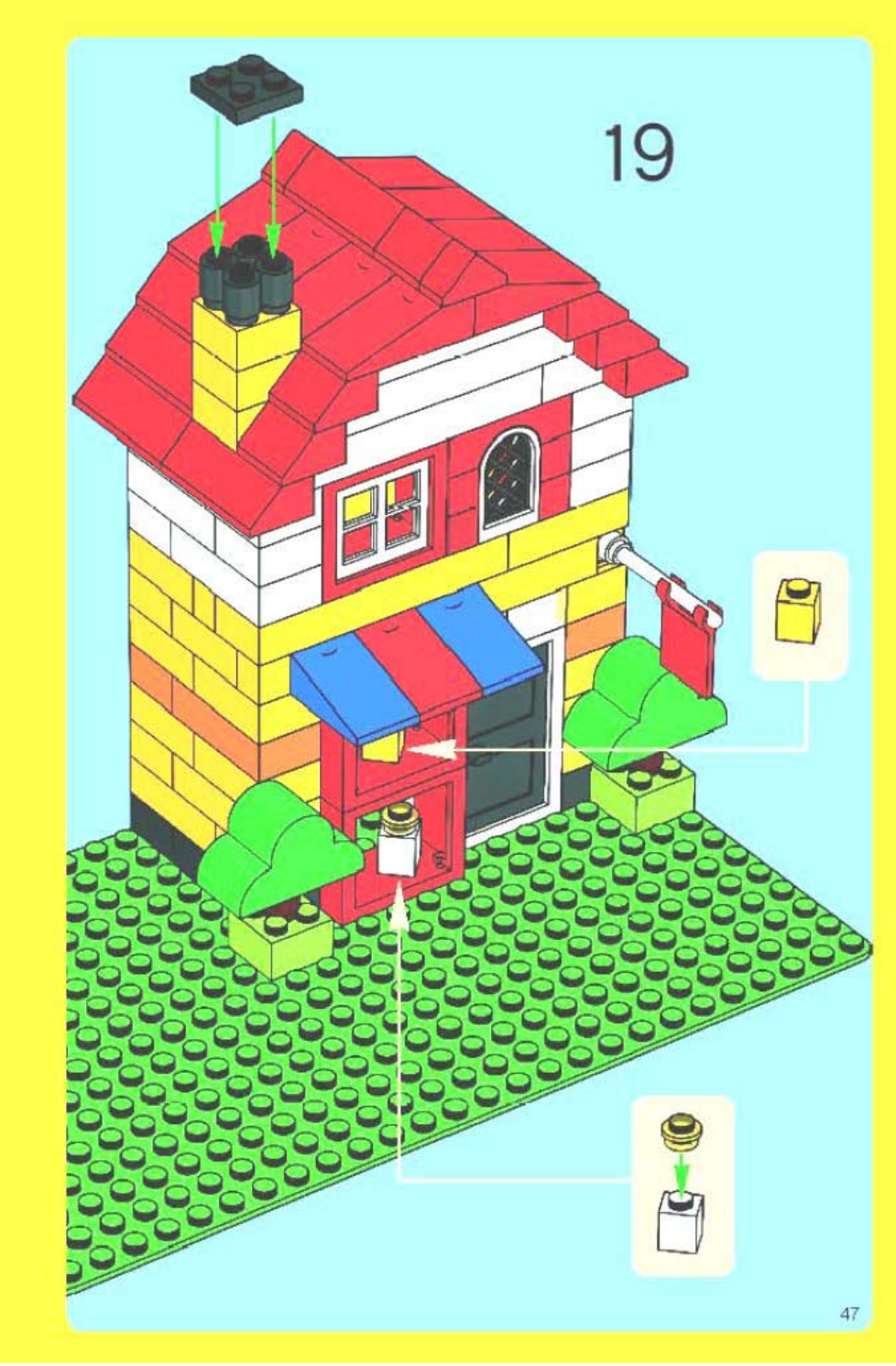

14
42


15
43


16
44


17
45

18
46


19
47