University of Oklahoma

GRADUATE COLLEGE

THE EFFECTS OF QUESTION CONTEXT AND RESPONSE FORMAT ON ACHIEVEMENT IN AN AVIATION MECHANIC WEIGHT AND BALANCE ASSESSMENT

A Dissertation

Submitted to the Graduate Faculty

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

Linda McCoy

Norman, Oklahoma

2003

UMI Number: 3085708

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE EFFECTS OF QUESTION CONTEXT AND RESPONSE FORMAT
ON ACHIEVEMENT IN AN AVIATION MECHANIC
WEIGHT AND BALANCE ASSESSMENT

A Dissertation APPROVED FOR THE
DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

BY

Raymond B. Miller (Chair)

Teresa K. DeBacker

Barbara A. Greene

Douglas F. Kauffman

A. B. Schwarzkopf

Acknowledgements

I have many people to thank for reaching this point. To my chair, Raymond Miller, thank you for helping me through the difficult writing process. It was a pleasure working with you. Thanks to members of my committee, Barbara Greene, Teresa DeBacker, Doug Kaufman, and Al Schwarzkopf for sharing their knowledge and experience in making this dissertation possible. I appreciate each of you for always giving me positive feedback.

I am deeply indebted to Ann Wilson, program analyst at the FAA, who wrote the Visual Basic program used to administer the experimental quiz. Ann's contributions made this research easier to collect and analyze. Terry Hollars and Johnny Malone provided excellent feedback in the development of the test instrument. I received exceptional support and encouragement from the following FAA supervisors and managers: Robert Kopecky, Gary Walker, Eldon Beavers, Marty Weaver, Dustin Sloan, Joe Tintera, and Bill Pickelseimer. I could not have completed this project without their help.

I am also indebted to the Part 147 schools, instructors, and administrators that allowed me to conduct research from their aviation students: Michael Highland and Rex Nivers, of Tulsa Technology Center; Edsel Ford, Canadian Valley Vo-Tech; Peter Lee, Robert Hensley, Greg Garcia, John Logan, and Ann Riley, Metro Tech Career Aviation Center; Fred Ritchey and Sean Fortier, LeTourneau University.

Finally, I want to thank to my family and friends for their patience and encouragement, especially my husband Don who kept me focused on reaching my goal.

TABLE OF CONTENTS

List of Tables

Table	Page

List of Illustrations

Abstract

This study investigated the effects of question context (word problems vs. abbreviated list) and response format (multiple-choice vs. constructed-response) on achievement in an aviation weight and balance quiz. An alternative question format was presented to simulate a more realistic problem situation. An experimental, computer-administered quiz was designed to measure the differences in test performance on the subject of weight and balance. Typical quiz items required the student to compute empty weight center of gravity. Other quiz items required knowledge of Federal Aviation Administration (FAA) rules regarding weight and balance measurements. A self-efficacy questionnaire was administered prior to the quiz. Distractors for the multiple-choice format were based on common misconceptions in interpreting FAA regulations.

A sample of 100 students from four aviation mechanic schools in Oklahoma and Texas participated in the experimental study. Scores for the multiple-choice questions were significantly higher than for the constructed-response questions. There was no difference in question context. Results suggested that (a) the multiple-choice format was easier than the constructed-response format, (b) students may not be learning procedures for solving real weight and balance problems, and (c) math computational skills were often weak. Analysis of student notes revealed no major differences in problem representation or solution between the question contexts and response formats.

Chapter 1

Introduction

Purpose

The purpose of this study was to investigate the effects of question context and response format on achievement in an aviation mechanic weight and balance quiz. Aviation mechanics must be knowledgeable in many diverse and complex study areas. Some example tasks include weighing aircraft, balancing loads on aircraft, fueling aircraft, reading blueprints and schematic diagrams, among others. The focus of my study is on the weight and balance problems. These problems require computations to adjust loads in different areas of the airplane to conform to engineering requirements for safe and efficient flight. As such, weight-and-balance problems can be viewed as a specific class of mathematical word problems. Mastery of these problems is critical for obtaining certification, which is currently evaluated with multiple-choice questions.

The present study looked at the impact of presenting students with weight and balance problems in an alternative question context (i.e., abbreviated list) and with an alternative response format (constructed-response). The constructed-response format requires the examinee to generate a response rather than select from a list of options (Bennett, 1993). The alternative formats were proposed to simulate a more realistic evaluation tool. Real-world problems are rarely presented as multiple-choice problems (Ryan and Greguras, 1998), nor are weight and balance problems presented in carefully constructed sentences. Instead, the aviation mechanic consults a certificate data sheet that contains concise, abbreviated lists of conditions and limitations under which the aircraft meets airworthiness requirements. The aviation mechanic then selects the relevant information, applies the appropriate rule or regulation, and computes a satisfactory answer.

The general research question guiding my research was: Is there a difference in cognitive test performance when participants use either multiple-choice or constructed-response questions, or when the item type is a story problem or an abbreviated list?

Problem Statement

Valid and reliable measurement of higher order knowledge is the subject of repeated research. How to accomplish this lofty goal in a specific subject domain was the focus of this study. One basic principle should underlie all measurement activities: the instrument or technique should fit the objective to be measured (Standards, 1985). The Standards (1985) describe validity as "the most important consideration in test evaluation" (p. 9). A second principle of equal importance is that no technique is worth using unless the results are reliable (Thorndike, 1971). An evaluation research study is useful only to the extent that it yields accurate, valid, and reliable data.

There have been many studies comparing test item format -- multiple-choice versus constructed-response – for psychometric trait equivalence, for cognitive information-processing research, for diagnostic value, in many different subject domains (Barnett-Foster, 1993; Martinez, 1999). There is little agreement between experts on the test format that is most appropriate for measuring mathematical ability. Researchers agree that the knowledge domain is a crucial determinant for selecting the most effective format for measurement purposes (Barnett-Foster, 1993; Traub, 1993).

The widespread use of multiple-choice test items for measuring higher order knowledge is perceived by many assessment experts to be a threat to test validity (Frederiksen, 1984), although the scores often possess high accuracy and reliability. The debate concerning which method of assessment is superior – multiple-choice versus constructed-response test items – has been often discussed, reported, and debated; yet no clear answer has emerged. Martinez suggested that a "mixture of item formats may yield the best possible combined effect" (p. 216). Rather than arguing which format is superior, combining multiple-choice with constructed-response questions on a test could take advantage of their positive features and minimize their liabilities (Martinez, 1999; Messick, 1993).

In addition, the present study also addressed the effect of varying question context, such as an abbreviated list versus a word problem. For example, would the student achieve the same result if the question were presented as a decontextualized mathematical expression as a replacement for a highly contextualized word problem?

Significance of Study

The present study contributes to the knowledge of testing behavior by providing data from a novel source. There have been no published studies in the areas of format and context effects related to aviation mechanics testing. The results should shed light on the feasibility of modifying the aviation mechanics test.

Summary

Chapter 1 has provided the setting of the proposed investigation along with the research questions. Chapter 2 reviews the literature pertinent to the research investigation. Chapter 3 outlines the method of data collection, the instruments used to evaluate the research questions, and the research design. Chapter 4 reports the results of the data analysis. Chapter 5 discusses the significance of the findings, limitations of the study, and suggestions for further research.

Chapter 2

Review Of Literature

Overview

The purpose of this chapter is to summarize the major research on mathematical word problem solving and response format. Several studies have explored the effects of rewording mathematical word problems into an easier to understand format (Cummins, Kintsch, Reusser, & Weimer, 1988; Davis-Dorsey, Ross, & Morrison, 1991; DeCorte, Verschaffel, & De Win, 1985; Reusser, 1990). Other studies have described mathematical problem solving as the efficient organization of knowledge (Mayer, Larkin, & Kadane, 1984; Sebrechts, Enright, Bennett, & Martin, 1996). Research conducted on multiple-choice and constructed-response formats has focused primarily on trait equivalence (Bennett, Rock, & Wang, 1991; Bridgeman, 1992; Frederiksen, 1984; Traub & Fisher, 1977; Ward, 1982), or item difficulty (Birenbaum and Tatsuoka, 1987; Katz, Bennett, & Berger, 2001; Ward, Frederiksen, & Carlson, 1980). Some research has been conducted on the roles of self-efficacy and item format with respect to mathematical problem solving (Pajares & Miller, 1997; Hackett, Betz, O'Halloran, & Romac, 1990).

Theoretical Foundations

When viewing math problems from a psychological perspective, there are many perspectives one could take. Greeno (1978) suggested that mathematical problem solving could be easier to understand as a production system. Productions take the form of if-then rules that specify when cognitive actions will take place. Productions constitute knowledge about "how to do things" (Anderson, 1983, p. 215).

Productions combine to form sets or systems. A production system represents knowledge of a complex procedure. Productions within a set are interrelated. A fundamental feature of a production system is that it contains a goal-subgoal hierarchy that interrelates all productions (Gagne, 1993).

When a mathematical computation is represented as a set of productions, each component of the skill is a production rule, consisting of a condition and an action. When the production system is executed, the conditions are tested, and if one is true, the action of the production is performed. A typical production rule in the ACT* (Adaptive Control of Thought) theory is represented by the addition of two numbers: "IF the goal is to find the sum of n1 and n2, and n1 + n2 = n3, THEN say n3"(Anderson, 1992, p. 167).

This general production rule has a specific goal, has variables to place specific numbers (n1, n2, and n3), and requires retrieval of a specific sum from long-term memory. It can be transformed into a specific rule that involves only declarative knowledge, "IF the goal is to find the sum of 6 and 5 THEN say 11" (Anderson, 1992, p. 167). The problem solution involves identifying relevant features of the problem, and performing the necessary transformations, which could involve adding new components or relations to the situation, or changing some part of the situation (Greeno, 1978).

Another perspective on solving mathematical computations is the relationship to skill in reading comprehension (Kintsch & van Dijk, 1978; Kintsch & Greeno, 1985). The concept of language comprehension refers to the construction of a meaningful representation of some information. Mathematical word problems are often difficult for students to solve. When solving word problems, students must move between different linguistic and symbolic codes (Wyndhamn & Saljo, 1997). Why word problems are so difficult could be related to a deficiency in either the student's mathematical knowledge or language comprehension skills (Cummins et al., 1988; De Corte et al., 1985; Kintsch, 1998; Reusser, 1990).

Reading Comprehension

The processing strategies for solving mathematical word problems can be related to general issues of prose comprehension (Hinsley, Hayes, & Simon, 1977; Kintsch & van Dijk, 1978). Understanding a problem text requires constructing a problem representation of the text in which problem-solving processes can operate (Kintsch & Greeno, 1985). Kintsch and Greeno (1985) proposed a dual representation, a propositional text base and a situation model, in order to comprehend word problems. The student is tasked with constructing from the verbal form of the problem a conceptual relation among quantities that guides the choice of calculations to be performed (Kintsch and Greeno, 1985, p. 110).

The Kintsch and Greeno (1985) model combined a set of knowledge structures with a set of strategies for using the knowledge structures in building the representation and solving the problem. The model transformed verbal sentences into a list of propositions. Propositions form the basic unit of language, and a proposition is simply stated a "predicate-argument schema" (Kintsch, 1998, p. 37). The Kintsch and Greeno (1985) model includes three sets of knowledge structures used to represent and solve problems: (a) propositional sets, (b) general schemas to represent problems and relations of sets, and (c) strategic knowledge to perform mathematical operations. The strategies used to represent and solve problems in the Kintsch and Greeno (1985) model corresponded to those described in an earlier model of solving arithmetic word problems developed by Riley, Greeno, & Hinsley (1983).

Word problems have received significant attention in the reading comprehension literature. Several studies have shown that rewording problems to remove ambiguous language can improve student's abilities to solve problems (Cummins et al., 1988; De Corte et al., 1985; Reusser, 1990). De Corte et al. investigated the influence of changes in wording of simple arithmetic problems without affecting the semantic structure or the level of difficulty of those problems. Their research produced evidence that the semantic structure of verbal problems strongly influenced the difficulty of solving these problems for first and second graders. Results showed that students failed to solve problems because they were unable to represent the problem correctly.

The text comprehension process was further explored by Cummins et al. (1988). In the Cummins et al. (1988) study, one type of arithmetic problem was solved by all first grade children when presented in numeric format, but by only 29 percent of the children when presented as a word problem. The relative ease in solving problems presented in numeric format suggested that semantic structure was more a factor than mathematical skill in problem solving. Cummins et al. proposed that word problems required a mapping onto the reader's knowledge base. Recall and solution errors were predicted to vary systematically in that solution errors would result from miscomprehended problems.

The impact of both personalizing and rewording mathematical word problems was tested in yet another study by Davis-Dorsey et al. (1991). The reworded problems simplified language of conventional textbook problems. The intent was to tap into existing schemata and make the interrelations of problem elements more specific. Results showed that less experienced problem solvers benefited from personalization and rewording together. More experienced problem solvers benefited from personalization but not from rewording. Davis-Dorsey et al. (1991) concluded that older students had better developed schemas than younger students for recognizing problem templates and were better able to understand the propositions required by each (p. 67).

Several researchers (Cummins et al., 1988; Greeno, 1978; Reusser, 1990) maintain that individuals acquire mathematical knowledge by solving problems in concrete situations. Reusser suggested that translating from a propositional text base into a mathematical problem is particularly difficult for learners who lack specialized arithmetic strategies or powerful verbal reasoning strategies (p. 4). Reusser used a model that focused on situational factors or world knowledge, in addition to linguistic knowledge, to facilitate understanding between the propositional text base and the arithmetic problem model. Results supported the hypothesis that using an elaborated problem text facilitated problem solving in that learners are able to construct a more appropriate problem representation.

The reading comprehension literature also includes studies in which rewording problems into a simpler linguistic structure did not yield such positive results. Other studies (Moyer, Sowder, Threadgill-Sowder, & Moyer, 1984;

Threadgill-Sowder, Sowder, Moyer & Moyer, 1985) contradict the use of a simplified language format in solving mathematical problems. They found that using a telegraphic list format in lieu of the typical sentence format did not make problem solving easier for students with low reading comprehension skills. Apparently, the low verbiage format did not offer enough context to make the problems understandable. Threadgill-Sowder et al. also found that using a drawn format was helpful for the low comprehension group.

In the current study, I examined whether simplifying the reading comprehension demands of the weight-and-balance word problems presented to aviation mechanics students improved their problem solving effectiveness. If the problem were presented in a different format, such as an abbreviated list in which extraneous verbiage were removed from the text, then it would be conceivable that students would be able to extract the salient information from the problem. They would rely on their mathematical knowledge rather than their linguistic knowledge to solve the problem. I analyzed student notes to see how students represented the problem solution – whether they constructed a drawing or table that represented an attempt to organize the elements into a recognizable problem structure. For example, a typical weight and balance ***word*** problem would resemble the following:

> As weighed, the total empty weight of an aircraft is 5,995 pounds with a moment of 885,997 pound-inches. However, when the aircraft was weighed, 24 pounds of potable water were on board at 86 inches, and 18 pounds of hydraulic fluid were in a tank located at 125 inches. What is the empty weight center of gravity of this aircraft?

The same problem, when rewritten into an ***abbreviated list*** format, would remove all linguistic structures as shown below:

Given:
Aircraft EW = 5,995 lbs
Moment = 885,997
Potable water = 24 lbs @ +86
Hydraulic fluid = 18 lbs @ +125
EWCG = ?

Changing from a semantic to a symbolic structure should make the problem solving process easier and more realistic for the aviation mechanic who is used to reading aviation type data sheets in which aircraft information is typically presented. See Appendix A for an explanation of weight and balance theory as it applies to aviation mechanics.

As neither the production model nor reading comprehension fully explains one's ability to solve math computations, a third perspective is that word problem solving is the successful combination of both reading comprehension and procedural knowledge. The following paragraphs propose that mathematical ability requires effective reading comprehension skills and relevant productions working together.

Mathematical Ability

Mathematical ability refers to the ability to solve mathematics problems efficiently and can be measured through cognitive analysis (Mayer, Larkin, & Kadane, 1984). Measurement of mathematical ability is a key component in tests of intelligence, achievement and aptitude. Studies comparing expert-novice problem solving performance suggest that knowledge, specifically domain-specific knowledge that is organized efficiently, is basic to problem solving performance (Chi, Glaser, & Rees, 1982; Larkin, 1983). Theories of problem solving are based on the idea that solving a particular problem requires domain-specific knowledge and general strategies (Greeno, 1978, Mayer et al., 1984). Research studies have found that intelligence, mathematics achievement, reading ability, and general reasoning ability seem to be related to student's problem solving ability (Latterell, 2000).

Mathematical problem solving is comprised of four phases – translation, understanding, planning, and execution (Mayer et al., 1984; Sebrechts et al., 1996). Each phase requires different knowledge. Linguistic and declarative knowledge is required for the translation phase, while schematic knowledge is necessary for understanding. Strategic knowledge, or how to solve for X, is required for planning. Algorithmic knowledge, such as the procedure for finding "3 x 5 = ___" is required for the execution phase. Individual differences may be due to differences in quantity and quality of each of these four kinds of knowledge (Mayer et al., 1984).

A person needs to retrieve knowledge of specific mathematical relations from long-term memory to recognize problem forms such as "total = time x rate" (Sebrechts, Enright, Bennett, & Martin, 1996). Errors in problem solving may occur when a person miscategorizes the problem or uses an inappropriate schema. Experts are able to categorize problems after hearing only a few words of the problem statement (VanLehn, 1989). Successful problem solvers are able to categorize mathematics problems on the basis of structural rather than surface features (Sebrechts et al., 1996).

Strategic knowledge involves knowing how to set goals and which procedures are effective in reaching these goals. The strategy that a person uses for a particular problem may be influenced by the format of the problem; different problem representations may lead to qualitatively different solution strategies. Mayer (1978) provided subjects with premises in a meaningful or organized format and found different strategies were used to solve problems when compared to subjects given premises in fragmented or nonsense format. The finding was replicated comparing problems presented in an equation format versus word format (Mayer, 1982). The group that received the equation format was significantly faster at solving the problems than the group that received the word format (Mayer, 1982). Mayer found that students presented with the equation format tended to use an isolate strategy while the students presented with the word format tended to use a reduce strategy. Both strategies are different methods of solving equations. If the equation is provided, the student attempts to isolate the unknown.

Algorithmic knowledge refers to knowledge about how to do something – how to carry out a procedure. Examples include arithmetic procedures such as knowing how to perform long division or three column subtraction, and algebraic procedures such as how to divide both sides of an equality by the same number (Mayer et al., 1984). An algorithm is an exact procedure for carrying out some task, such as adding two numbers. One component of a person's algorithmic knowledge involves arithmetic algorithms. Simple procedures can be solved by rote. More complex procedures require a series of steps in which the individual has several chances to make errors. Errors in execution of a procedure may account for many transcription problems. For example, a decimal point may be misplaced or a value changed inadvertently (Sebrechts et al., 1996). Multiple-choice questions reduce these kinds of procedural errors because the answer is constrained from the available alternatives.

In the present study of aviation weight and balance problems, a cognitive analysis of each question was conducted to validate the mathematical problem-solving model used by Mayer et al. (1984). The questions were evaluated to determine which, if any, of the four phases of mathematical problem solving --translation, understanding, planning, and execution--were required to solve the questions. See Appendix B for a step-by-step analysis of each question.

Regardless of which perspective best explains how an individual arrives at the problem solution, measurement of mathematical ability remains a controversial subject. The following section presents alternative views of assessment of mathematical ability through multiple-choice or constructed-response questions.

Comparison of Response Format

A recent review of the response format literature (Martinez, 1999) addressed concerns about the different cognitive abilities that multiple-choice and constructed-response item formats are able to measure. Multiple-choice format has the benefit of broad domain sampling, test and scoring reliability, and economy, while constructed-response format is better suited for complex cognition (Martinez, 1999, p. 216).

Multiple-choice testing is used more often than other test formats because it is objective, less costly to administer and to score, and has excellent reliability (Barnett-Foster, 1993; Ebel & Frisbie, 1991). Sampling a broad range of topics with a few multiple-choice questions is a highly desirable feature for large-scale applications, such as advanced placement (AP) subject area examinations or airman knowledge certification tests. The multiple-choice test format is also useful in the mathematics and science disciplines because evaluation tends to focus on a single correct answer

(Barnett-Foster, 1993; Braswell & Kupin, 1993; Campbell, 1999). This quality makes the multiple-choice format a very efficient method of assessment.

On the other hand, multiple-choice tests have been criticized by many prominent educators for dealing with only a small portion of mathematical problem solving (Latterell, 2000). Criticism of mathematics tests has focused on the predominance of multiple-choice test items requiring factual knowledge rather than items that assess cognitive abilities, such as reasoning, critical thinking, problem solving, interpreting and applying ideas (Collis & Romberg, 1991). According to the National Council for Teachers of Mathematics' (NCTM) published standards for assessment, tests should assess student ability to use mathematics to solve problems.

Although some researchers (Boodoo, 1993; Frederiksen, 1984; Latterell, 2000; and Snow, 1993) have criticized multiple-choice questions for measuring primarily factual knowledge, others have argued that it is possible to measure complex thought (Aiken, 1982; Ebel & Frisbie, 1991; Gronlund, 1982; Haladyna, 1997). Researchers agree that it is difficult to construct multiple-choice items that encourage generation of novel applications or original interpretation (Ward, Frederiksen, & Carlson, 1980).

An important limitation of multiple-choice test items is their inability to diagnose student errors unless distractors represent common misconceptions (Birenbaum & Tatsuoka, 1987; Latterell, 2000). Another criticism of multiple-choice items is the relative ease in guessing the correct answer (Barnett-Foster, 1993). Through a process of elimination, examinees may deduce the correct answer. The research on guessing behavior indicates that examinees use partial knowledge to reduce the number of choices to a subset of choices from which they can randomly guess (Barnett-Foster, 1993). As a result, examinees receive full credit for answers they do not know (Barnett-Foster, 1993), which could be an advantage in mathematical computation items because students are able to recognize minor arithmetical mistakes and are not penalized for a computation error (Braswell & Kupin, 1993).

Constructed-response items require the student to generate the correct answer, not just to recognize it (Bridgeman, 1992; Katz, Bennett, & Berger, 2001). Multiple-choice items create a situation in which the reader focuses on finding the right answer rather than constructing some meaning from the question (Campbell, 1999). Constructed-response items allow for more inferences about the thought processes contributing to the answer (Collis & Romberg, 1991). However, constructed-response items are seldom administered because of the following problems: complexity, time needed to construct a response, sample selection, and test reliability (Ebel & Frisbie, 1991). Because constructed-response items are scored by humans rather than computers, interrater judgment is a valid concern (Collis & Romberg, 1991).

Constructed-response exam formats have many problems that detract from their ability to measure cognitive knowledge. They have been criticized for subjectivity and expense of scoring, decreased reliability, and narrow sampling range (Bennett, Rock, & Wang, 1991; Martinez, 1999). Constructed-response items are susceptible to bluffing (Latterell, 2000). The ability to write and express thoughts well has often resulted in higher scores than deserved (Latterell, 2000).

The comparison of multiple-choice and constructed-response test formats for trait equivalence has been the subject of many investigations. Although constructed-response items are expensive to score and require more time from the examinee to answer, they are believed to measure traits that cannot be measured with multiple-choice items (Frederiksen, 1984; Birenbaum & Tatsuoka, 1987).

Frederiksen (1984) reported that testing increases retention of material and that short answer or completion items may be conducive to long-term retention. Content differences are more important than format differences in improving performance and retention of material. Teaching for the test is commonplace in many aviation schools and in typical college-entrance preparatory courses. Frederiksen feared that teachers were switching to the multiple-choice tests for accountability reasons, rather than testing for abilities that are not easily measured by multiple-choice items.

> "First, format may influence the test developer's selection of the task to be posed by each item. Those choices, as well as the format itself, may in turn influence how and what teachers teach and how students prepare for a test. More fundamentally, format may influence the cognitive processes involved in dealing with test items and hence the nature of the skills taught and learned" (Frederiksen, 1984, p. 195).

Critics of multiple-choice items claim that only rote facts can be tested in this way, not deep understanding (Snow, 1993; Traub, 1993). Many well-known ability and intelligence tests confirm this criticism. The GRE Advanced Psychology Test was found to measure primarily factual knowledge (Frederiksen, 1984). Memory items comprised 70 percent of the test. The Orthopedic In-Training Examination, which was designed to measure competence in orthopedic medicine, was also found to measure primarily recall of information rather than interpretation of data, application, or evaluation (Frederiksen, 1984).

Conversely, Haladyna (1997) described how to measure understanding, evaluation, prediction, and problem solving with the multiple-choice format. Test items that measure understanding require students to identify a fact, concept, principle, or procedure. Students select the correct definition from a list of choices, or identify characteristics or examples from nonexamples. With critical thinking items, students may be asked to predict what will happen or to select and use a criterion.

Recent studies reaffirm the controversy concerning which response format is more difficult. Latterel (2000) reached the conclusion that "In general, research suggests that multiple-choice mathematics questions are more difficult for students than equivalent open-ended mathematics questions." (p. 29) Another study of mathematical word problems on the Scholastic Aptitude Test reported that "Researchers have frequently noted that some items are more difficult in the constructed-response format than in the multiple-choice format" (Katz, Bennett, and Berger, 2001, p. 39). Katz, Bennett, and Berger (2001) hypothesized that differences in performance between constructed-response and multiple-choice formats is the result of using different strategies in solving the problems. They concluded that comprehension difficulties were more likely to explain performance differences than choice of strategy (Katz, Bennett, and Berger, 2001).

The range of cognitive demands appears to be domain-related (Barnett-Foster, 1993; Traub, 1993). The majority of studies of verbal ability suggest that multiple-choice items measure different traits than constructed-response items (Ackerman &

Smith, 1988; Traub & Fisher, 1977; Ward et al., 1987). Multiple-choice measures of writing ability do not accommodate some cognitive skills, such as originality and creativity (Ackerman & Smith, 1982); but for tests of quantitative ability, no apparent format effects were found (Barnett-Foster, 1993; Bennett et al., 1991; Traub and Fisher, 1977; Sebrechts et al., 1996).

Response format did not reveal a significant difference in studies conducted by Traub and Fisher (1977) and by Ward (1982) because the constructed-response items were revised versions of multiple-choice items. Traub and Fisher tested eighth grade students with parallel constructed-response and multiple-choice forms of a verbal comprehension test. The students were also tested on mathematical reasoning, recall memory, recognition memory, and predisposition to guess answers to multiple-choice items. The constructed-response items were administered first to eliminate learning from multiple-choice cues, and the multiple-choice test was administered two weeks later. The mathematical reasoning tests measured the same attribute, regardless of test format. Traub and Fisher reported a significant effect on the verbal comprehension test, but it was unrelated to recall memory, recognition memory, or guessing. No significant effects were reported for the mathematical reasoning tests.

Ward (1982) examined the constructed-response and multiple-choice formats used for the GRE verbal aptitude test. Antonyms, sentence completion, and analogies were selected for relative ease in transforming into the constructed-response format, for producing reliable, easy to score results, and possibly for measuring complex ideas. The items were presented in four different ways – standard multiple-choice, single answer, multiple answer, and a list format. Ward examined the data in a two-factor structure – item type and item format. An exploratory factor analysis was used to examine the influence of response format. From the principal axis factor analysis, he concluded there was no format difference in the antonym and sentence completion items. The analogy tests revealed a small second factor (only five percent of the common factor variance). Ward interpreted the result as a speed factor because the analogy test was the only test that was at all speeded.

When comparing his study to an earlier problem-solving study by Frederiksen (1980), Ward (1982) admitted that generating a single response did not compare to reading and comprehending passages containing a number of items of information relevant to a problem and writing several answers. The problem-solving studies required determining the relevance of information to apply reasoning and inference to draw conclusions, which required specialized knowledge. Subjects were asked to compose responses relating several complex ideas to one another, thus constituting "ill-structured" problems. In contrast, the verbal aptitude items of Ward's study constituted "well-structured" problems.

When Ward et al. (1980) reversed the process (constructed-response items were changed to multiple-choice format), there were significant format differences. Ward et al. simulated a problem that showed a research study, graph with results, and statement of the major finding. The task was to write possible explanations or hypotheses that could account for the finding. The multiple-choice form listed hypotheses based on answers from the constructed-response test. Scores reflected quality, number, and unusualness of hypotheses. The multiple-choice format produced the highest achievement scores. As expected, students were able to discriminate good from poor alternatives even when they could not generate a good response (Ward et al, 1980, p. 18). Correlations between corresponding scores for the two formats were low. Both correlated with the knowledge factor and verbal and reasoning factors on the GRE Advanced Psychology Test. Only the constructed-response items correlated with ideational fluency, which was interpreted as searching long-term memory for relevant ideas (Ward et al., 1980).

A study conducted by Bennett, Rock, and Wang (1991) used the 1988 Advanced Placement Computer Science (APCS) examination, which combined multiple-choice and constructed-response items. The design paralleled an earlier study in that it examined a two-factor model using maximum likelihood factor estimation. The authors did not find significant differences between multiple-choice and constructed-response items. The constructed-response items required the student to decompose the specification into goals, formulate plans to achieve each goal, translate the plan into Pascal code, and debug the code by mentally simulating its effect. The authors speculated that while one multiple-choice item could not cover the depth of one constructed-response item, a combination of 50 such items could tap some of the same processes. Their analysis suggested that the multiple-choice items overlapped the constructed-response items in some processes.

Several limitations of the Bennett et al. (1991) study were noted. The constructed-response items did not represent the length or the complexity of real world programming problems, and the scoring method combined several dimensions, including one score, and did not account for others, such as originality. The assessment did not address diagnostic errors. The authors concluded that constructed-response items provide a trace to the examinee's solution process, which multiple-choice items can not duplicate.

Bridgeman (1992) conducted a study on the GRE-Quantitative (GRE-Q) test that compared multiple-choice and constructed-response formats in both a scannable format and a computer-administered mode. He addressed the questions of format equivalence in terms of difficulty, discrimination, and correlational structure. Students were able to input numerical answers with either a gridded sheet that accommodated decimals, fractions, negative numbers, and equations with one variable or the computer keyboard.

Bridgeman found that there were differences between the two formats. Some items that were relatively easy in the multiple-choice format were relatively difficult in the constructed-response format. Format effects were found when the multiple-choice options were not an accurate reflection of the errors actually made by students. Bridgeman concluded that the constructed-response format was superior for describing the specific skills of the student and for eliminating random guessing. However, he also found that total test scores for the two formats were comparable.

Birenbaum and Tatsuoka (1987) examined the effect of item response format on the diagnosis of examinee misconceptions in fraction-addition arithmetic operations. The subjects were 285 eighth grade students from a mathematics lab. The multiple-choice responses were constructed from common errors on the constructed-response pilot test. The study was designed to test equivalence of the two formats. They conducted three analyses to examine the effect of the response format. The first analysis focused on the underlying structure of the two test forms. The second analysis compared the two formats with respect to the type of errors committed, and the third analysis focused on diagnosing the source of student misconceptions with respect to fraction-addition operations. The authors used multidimensional scaling to map the test items into a two-dimensional space. The constructed-response items formed two distinct clusters, for items with like denominators, and one for unlike denominators. The multiple-choice items showed no clear distinction between types of items. Basic test characteristics showed the tests were homogeneous and reliable.

Additional analyses were used to study types of errors and to diagnose sources of misconceptions. The results of the error analysis showed that significantly larger number of error types occurred in the multiple-choice items. The authors inferred that students were less consistent in applying the rules of operation for solving procedural tasks with the multiple-choice format, and that the cognitive processes involved in the two formats were different. The constructed-response items required computing the answer without cues or distractors, while the multiple-choice test items allowed retrieval of answers from the distractors. Student effort was directed toward judging the answers as correct rather than carrying out the calculations. Since the distractors represented common errors rather than random incorrect errors, the process was more complicated and resulted in more error types. Birenbaum and Tatsuoka concluded that constructed-response items were more appropriate for diagnosing student errors and for evaluation of procedural tasks.

Sebrechts et al. (1996) evaluated algebra word problems in the constructed-response format that were previously administered to college students as multiple-choice items on the Graduate Record Examination (GRE). Performance scores in the

Sebrechts et al. study were correlated with standard measures of problem difficulty and student proficiency as those given on the GRE multiple-choice exams. Scores also correlated with Scholastic Aptitude Test (SAT) mathematics scores (Latterell, 2000).

Several studies reported differences on mathematical computations where response format was varied (Bridgeman, 1992; Traub, 1993). Response time per question should be less on multiple-choice test questions than on constructed-response test questions, regardless of question context. Research has shown that recognition of the correct answer from a list of responses gives a distinct advantage to the multiple-choice test format (Lukhele, Thissen, and Wainer, 1994).

Self-Efficacy

Many research studies were flawed in that they did not compare parallel tests or that the differences in format were the result of another variable, such as self-efficacy (Latterel, 2000). Self-efficacy is defined "as the conviction that one can successfully execute the behavior required to produce the outcomes" (Bandura, 1977, p. 193). Pajares and Miller (1997) reported that student's judgments of their capabilities to solve mathematics problems were predictive of their actual capability to solve those problems. "Math self-efficacy has been shown to be as strong a predictor of mathematical problem-solving ability as general mental ability (Pajares & Miller, 1997, p. 214)."

Pajares and Miller (1997) studied the relationship between mathematics self-efficacy and item format. They sampled middle school students to determine if there is a difference in self-efficacy judgments and test format. Students were presented with varying formats of the self-efficacy instrument and the performance assessment. Pajares and Miller hypothesized a close link between the efficacy judgment and the criterion task. Students typically are overconfident about their ability to solve mathematics problems (Pajares & Miller, 1997). No significant differences in self-efficacy judgments were found between multiple-choice and constructed-response formats. Self-efficacy was included as a covariate in the present study to provide a control of mathematical problem solving ability.

Summary

The studies reviewed provide evidence of the complexity of measuring mathematical word problems. The reading comprehension literature provided a wealth of information concerning the mental processes that are present in solving word problems. Rewording problems to simplify the semantic structure had mixed results. Some data indicated that response format made a difference in assessment of higher cognition. However, other studies demonstrated that procedural knowledge could be measured adequately with multiple-choice questions. Multiple-choice test items are not necessarily easier, nor are they only measuring factual knowledge. In the current study, I attempted to identify which factor would affect test performance – question context or response format. The following research questions guided the process.

Research Questions

The general question of interest to the present study is as follows: Is there a difference in cognitive test performance when participants use either multiple-choice or constructed-response questions, or when the item type is a story problem or an abbreviated list?

The specific questions that this study addressed are the following:

1. Does response format (multiple-choice versus constructed-response) affect achievement on aviation weight and balance problems, controlling for self-efficacy?

2. Does question context (word story problem versus abbreviated list) affect achievement on aviation weight and balance problems, controlling for self-efficacy?

3. Does the interaction of response format and question context affect achievement on aviation weight and balance problems, controlling for self-efficacy?

In addition, I examined item analysis data (difficulties), item response times, and student notes and written computations to the items to see if there were any patterns that helped explain participant performance. In the next chapter I will present the details of the methodology of the present study.

Chapter 3

Methodology

Overview

The primary purpose of this study was to establish the relationship between question context and/or response format and achievement on a written test in a mathematical problem-solving domain. The following research question investigated the potential effects on achievement made by either question context or by response format: Is there a difference in cognitive test performance when participants use either multiple-choice or constructed-response questions, or when the item type is a story problem or an abbreviated list? Specifically,

1. Does response format (multiple-choice versus constructed-response) affect achievement on aviation weight and balance problems, controlling for self-efficacy?
2. Does question context (word story problem versus abbreviated list) affect achievement on aviation weight and balance problems, controlling for self-efficacy?
3. Does the interaction of response format and question context affect achievement on aviation weight and balance problems, controlling for self-efficacy?

This chapter describes the research design, participants, and sampling procedures used in this study. A thorough description of the development and revision of the experimental quiz is included, along with test specifications and cognitive question analysis.

Research Design

This study used a 2 x 2 repeated measures analysis of covariance (ANCOVA) research design. The between-subjects variable was question context (word story problem versus abbreviated list). The within-subjects variable was response format (multiple-choice versus constructed-response). Self-efficacy was used as a covariate since self-efficacy has been found to influence mathematical problem solving in some recent studies (Pajares and Miller, 1984; Latterel, 2000). To avoid an order effect, response format and questions were counterbalanced.

Participants

This experimental study was administered to 100 students enrolled in aviation maintenance technology schools in Oklahoma and Texas between February and June 2002. Cohen (1977) recommended power analysis to establish the minimum number of students per group. Group size was determined to be 44 to produce a medium effect size of .30, power of .80. Participants were randomly assigned to groups. Class size varied considerably between career tech schools and universities. Participants included 94 males and 6 females enrolled at four different aviation schools. Participants were categorized into five separate age ranges (Table 1a). Most of the participants were adults, however, there were a few high school students included in the study (Table 1b).

Students had completed the following core subject areas in the aviation mechanic general curriculum: mathematics, physics, and weight and balance. Table 1c shows the number of prior mathematics courses taken by participants. Nearly half the participants reported no prior aviation experience (n = 46). Another 11 reported military experience, and 42 reported civilian aviation experience.

The majority of participants reported English as their native language (n = 78). Approximately 80 percent reported their English proficiency as Good, Very Good, or Excellent. (See Table 1d.) The computer program malfunctioned at one school and did not save answers to the English proficiency question. I determined that this information was incidental to the study and did not affect the outcome of the primary research questions. A few participants did not answer any questions on the student information screen; thus, no data are available for some background questions.

Instrument

Test development. The FAA aviation mechanic general certification test bank was the main source of questions for this study. Approximately ten questions were selected from the weight and balance, physics, and mathematics sections of the general exam. Each question required a mathematical computation presented as a word problem with multiple-choice response format. Average p-values (ease index) and response times of the selected test questions were computed using data from tests administered between January 1 and September 6, 2000. (See Table 2.)

Four of the questions selected for the quiz were rewritten into the constructed-response format. The constructed-response questions contain stem-equivalent mathematical operations to those in the multiple-choice questions. Slight modifications were made to two questions in the abbreviated list format in order to remove contextual information from the equation. Two additional questions from the

Table 1

Number of Participants by Age, Educational Level, Prior Mathematics Courses, and English Proficiency

a. Age		b. Educational Level	
Group	n	Group	n
15-17	6	High school student	9
18-21	33	High school graduate	25
22-25	25	GED equivalent	4
26-29	12	1-2 years college	54
30+	22	College graduate	6
N/A[a]	2	N/A	2

c. Prior Mathematics		d. English Proficiency	
Group	n	Group	n
Algebra	87	Excellent	29
Geometry	71	Very Good	24
Trigonometry	47	Good	27
Calculus	35	Fair	6
Physics	55	Very Poor	1
		N/A	13

[a]N/A = not available

same content area were constructed to balance the exam after the pilot test data were analyzed. The written quiz was converted into a Visual Basic 5.0 computer-administered application entitled AVQUIZ to simulate the same testing environment in which students take the FAA aviation mechanic general examination. This feature allows students immediate feedback on their scores and provides additional feedback to the test developer for analysis purposes

Table 2

Mean P-Values and Response Times in FAA Item Bank

Question #	P-Value	Response Time
1	.86	137 s
2	.77	176 s
3	.85	102 s
4	.78	199 s
5	.81	250 s
6	N/A[b]	N/A
7	.62	303 s
8	.70	417 s
9	.80	303 s
10	.84	301 s

Note. P-values and response times are based on multiple-choice questions in airman knowledge test bank.

[b]N/A = not available.

Technical review. Two subject matter experts (SMEs) reviewed AVQUIZ for accuracy. Both SMEs were certified aviation mechanics with test question writing responsibilities. One was the test writer for the aviation mechanic general test, and the other SME was the author of the Aircraft weight and balance handbook, FAA-H-8083-1, 1999. This handbook was the official FAA reference used at all Part 147 schools for teaching weight and balance theory and application at the time of this study.

Both SMEs completed the abbreviated list format and noted several technical errors. Two instructors at a nearby community college also provided technical advice. The instructors suggested test bank questions that were directly relevant to the job students were being trained to perform. After examining all technical review comments, the test developer replaced several questions on the quiz and made the following revisions.

- Question 3 – Invalid answer in computer score key was corrected.
- Questions 2 and 3 – Vocabulary was made consistent throughout the problems, and unit of measurement was changed from "inches" to "feet."
- Question 4 – Typographical error was corrected.
- Questions 5 & 6 – Clarified the nature of ballast (temporary or permanent) in the stem of the problem.
- Questions 7 & 8 – Clarified confusing wording in the stem.
- Questions 9 & 10 -- Removed questions from quiz due to content validity.

Test specifications. The test specifications used for development of the quiz are presented in Table 3. Two practice questions familiarize students with question and response format. The practice questions require simple arithmetical computations about area of geometric shapes and are relatively easy to solve. Questions were derived from the mathematics section of the general exam. Four questions assessed knowledge of Newton's third law of motion and volume of a three-dimensional object and were derived from the physics section of the general exam. The remaining eight questions required either (a) computation of general weight and balance problems, (b) knowledge of specific FAA rules regarding addition of fluids or ballast to the aircraft, or (c) computation of empty weight center of gravity after alterations have been accomplished to an aircraft.

Table 3

Test Specifications for Content Areas

Content Area	Word Problem	Abbreviated List
Area of Geometric Shapes (Practice)	1 - MC[c]	1 - MC
	1 – CR[d]	1 - CR
Newton's Law of Motion/Volume of Geometric Shapes	2 – MC	2 - MC
	2 – CR	2 - CR
Empty Weight Calculations	3 – MC	3 - MC
	3 – CR	3 - CR
Weight & Balance Alterations	1 – MC	1 - MC
	1 – CR	1 - CR

[c]MC = multiple choice

[d]CR = constructed response

Developmental Tryout

A developmental tryout was conducted in November 2000 at a midsize community college located in Tulsa, Oklahoma. Twenty-five adult students enrolled in the aviation mechanic curriculum participated. The adult instructor also completed the quiz and provided professional observations and feedback regarding the quiz. Before administering the quiz, the test developer briefed the class on the purpose of the quiz and gave verbal instructions on how the students were to complete the exercise.

The classroom computer lab had only seven working terminals, so the students rotated through the lab. As one terminal became available, another student would begin the exam, and so on. As the class had completed the weight and balance section three weeks prior to the quiz, many students commented they had already forgotten how to do some of the problems. The written questionnaire was subsequently removed from the testing protocol.

Participants. There were 23 males and 2 females in the class. Twelve students were between the ages of 18 and 21, five were between 22 and 25, three between 26 and 29, and four were 30 or older. Among the students, there were 11 high school graduates, one with a GED equivalent, 10 with one to two years' college, and one college graduate. The majority of students had completed algebra and geometry. Nineteen students had no prior aviation experience, while five had some civilian experience, and one was an employee of a major airline.

Results. Results from the quiz were low but not unexpected. The students knew the quiz did not count in their school grade or on the FAA test. While only a few students were able to solve the constructed-response questions, several students made common errors, either from an incorrect interpretation of the rule or from a numerical computation error. Several students simply omitted answering the constructed-response type of question, while no one omitted the multiple-choice questions.

Participant scores on the multiple-choice and constructed-response versions of the AVQUIZ were converted into percentages. Test scores for the abbreviated list group (M= 41.67, SD = 19.92, n = 12) exceeded those of the word problem group (M = 30.91, SD = 19.92, n = 11). Response times for the abbreviated list group (M = 16.14, SD = 9.67, n = 12) were considerably longer than response times for the word problem group (M = 9.85, SD = 8.29, n = 11). This finding supports the research literature that contextual information makes it easier for students to develop adequate problem representation (Choi and Hannafin, 1998; Kintsch, 1998).

Discussion. After all students had completed the quiz, a group discussion was conducted to allow students the opportunity to comment on the quiz. The instructor and several students pointed out discrepancies between the two versions. Question 1 concerning the placement of different boxes at various points in the aircraft was not identical between the two versions. The word problem contained three boxes, while the abbreviated list version contained four boxes, thus requiring one additional computation. The instructor also questioned a rule interpretation. This comment was provided to the SME responsible for the aviation mechanic general exam, and after agreement with the instructor's interpretation, the word "alcohol" was removed from the stems of two questions and replaced with the words "potable water."

Typographical errors were also noted on other questions. All discrepancies were corrected prior to the administration of the pilot study. See Appendix C.

Revised Quiz

The revised quiz was restructured into four test versions to remove any order effects. See Appendix D. Each of the four test versions consists of six multiple-choice questions, six constructed-response questions, and two practice questions (one multiple-choice and one constructed-response format). Versions 1 and 3 contain standard word problems, while Versions 2 and 4 contain abbreviated list problems. The same questions are presented in all four versions, but Versions 1 and 2 begin with the six multiple-choice questions followed by the six constructed-response questions. Versions 3 and 4 reverse the order and present the six constructed-response questions first, followed by the six multiple-choice questions. Students were randomly assigned to one of the four versions.

Procedure

All students submitted signed consent forms before participating in the study. High school students under the age of 18 also submitted parental consent forms. Refer to Appendix E.

Data for this study were collected from an experimental weight and balance quiz (AVQUIZ) administered by computer. Each student was given a written instruction sheet to minimize confusion at the start of the quiz (Table D1). Students provided demographic information by selecting their gender, age range, prior mathematics and science courses, prior military or civilian experience in the aviation mechanic job specialty area, and English proficiency. Table D2 contains the sample preliminary background questions.

Students completed a self-efficacy questionnaire prior to completing the 12-item quiz. Table D3 contains the prequiz questions administered to students who received versions 1 and 3 (word problems). Table D4 contains the prequiz questions administered to students who received versions 2 and 4 (abbreviated list). The questionnaire asked the students to rate their confidence level in answering sample weight and balance questions, using a five-point Likert scale. These questions mirrored the content and format of questions on the experimental quiz. The students were also provided two practice questions to familiarize themselves with the response format using the computer mouse and keypad.

Students were presented questions in a standard word problem format or an abbreviated list. Students were allowed the use of calculators and were provided blank notepaper during the quiz. All notepaper was collected after the quiz for further analysis. The Visual Basic 5.0 computer program AVQUIZ provided corrective feedback after each practice question. Message boxes prompted students to begin the quiz. Students were not allowed to return to questions. Students received a total test score but no confirmation for individual questions. Timing in seconds of each scored response was collected. Students were instructed to use the numeric keypad to enter answers for constructed-response questions and the computer mouse to select an answer for each multiple-choice question.

Each multiple-choice question consisted of one correct response and two incorrect responses. The incorrect responses were chosen from the most common errors produced by students on previous airman knowledge exams. These questions require recall and comprehension of relevant FAA rules, knowledge and selection of the appropriate mathematical algorithm, and performance of the mathematical operation.

Summary

An experimental weight and balance quiz entitled AVQUIZ was designed to simulate the testing environment that aviation mechanics experience during the certification process. Students were randomly assigned to one of four versions in which question context and response format were varied to measure differences in performance. The next chapter reports the results of the data collection and statistical analyses performed.

Chapter 4

Results

Overview

This chapter describes the data collection process, statistical analyses and qualitative analysis of results. The statistical analyses of the tests include descriptive statistics, reliability analysis, a 2 x 2 repeated measures analysis of covariance, and item analyses. Data analysis was accomplished to assess the extent to which inferences could be made regarding the validity and reliability of the test scores and appropriate conclusions could be made regarding the research questions.

Data Collection

Visual Basic 5.0 was used to collect data for the aviation mechanic quiz AVQUIZ. Data were saved on a disk at each computer workstation. Each workstation used the Windows 98 or Windows 2000 operating platform. AVQUIZ output included three separate Access 97 tables (Applicant, PreQuiz Results, and ExamResults), plus a tab-delimited ASCII text file.

Data from each school participating in the study were read into separate SPSS files and then combined into one master file for data analysis. Student answers to the four versions of AVQUIZ were recoded into one order in order to accomplish item analysis. Multiple-choice questions were coded 1 through 6, while constructed-response questions were coded 7 through 12. Because of scoring irregularities, data accuracy in the final SPSS file became a priority. Data were checked multiple times, with new scores computed and compared with the original scores assigned by the

Visual Basic program. The Access tables and ASCII text files provided a formal check and balance system in which coding errors were easy to identify.

Several program errors occurred during data collection. At one school, the program failed to collect answers regarding English proficiency at the student information screen. There were also problems with the multiple-choice scoring feature. Some questions and answers had been reordered after the pilot study, and the Visual Basic program scored those answers incorrectly. The researcher discovered this bug when a student complained. Eleven student records were manually checked for errors and scores were adjusted.

Other errors occurred with the constructed-response scoring feature throughout the data collection process. Instructions stated that answers should be recorded to the nearest hundredth; however, many students failed to observe this requirement. In one instance, the correct answer was "31.60" and the program did not recognize "31.6" as a correct answer. The researcher visually screened all constructed-response answers and manually rescored answers as correct if the student's answer fell within a hundredth of the preset answer key. The Visual Basic program did not have a built-in tolerance for rounding answers.

Although written instruction sheets were provided to each student before beginning the quiz, some students had problems with the computer disks and were allowed to restart the quiz. The problem occurred when the file path on the computer disk did not match the path named in the Visual Basic program. After the researcher discovered the source of the error, the computer disks were recopied with the correct file path. No further problems occurred with file saving after students began the quiz.

One student took the quiz twice. The researcher checked student identification numbers with consent forms and found one extra quiz had been recorded in the ASCII text files. The results from this second attempt were invalidated, removed from the Access database, and omitted in the final SPSS output.

Self-Efficacy Results

Pajares and Miller (1997) suggested that students' self-efficacy is better measured with questions similar to the actual task. Their research found that self-efficacy was not affected by presentation of the problems as either multiple-choice or constructed-response. They suggested that students in their study did not look at the alternatives when making their confidence judgments (p. 224). They also suggested that their students expected the problems to be in the traditional multiple-choice format and made their judgments accordingly (p. 224).

In this study, self-efficacy could have accounted for differences in exam performance between groups. Of the four schools represented, three were career tech programs and one was a private college. To check for differences in self-efficacy among the schools, the individual ratings for all four items were summed for each participant, with 20 being the highest possible score and 4 being the lowest. A one-way analysis of variance revealed a significant difference between groups [$F(1,3) = 24.09$, $p < .001$)]. Mean scores on the self-efficacy measure are shown in Table 4 for the four schools. Post hoc comparisons confirmed the college group differed significantly from the career tech groups on the confidence measure.

Table 4

Prequiz Self-Efficacy by School

Group	M	SD	n
Career Tech CV	11.00	3.16	10
College LT	17.86	2.38	35
Career Tech MT	12.27	3.89	44
Career Tech TT	12.82	2.64	11

Table 5 summarizes the number of students selecting 1 through 5 on the self-efficacy scale for each of the four prequiz items. Table 6 lists the Spearman rho correlation between the four prequiz self-efficacy items and their corresponding AVQUIZ multiple-choice and constructed-response items. Most students were confident that they could solve the sample weight and balance problems, regardless of question context.

Table 5

Prequiz Self-Efficacy Frequencies

	Response Category				
PreQuiz Item[e]	1	2	3	4	5
1. A rectangular shaped fuel tank measures	4	15	13	17	51
2. An aircraft as loaded weighs	8	18	34	16	24
3. As weighed, the total empty weight of an aircraft is	8	17	23	25	27
4. An aircraft with an empty weight of	8	17	27	19	29

[e]See Appendix D, Table D3 for complete text of prequiz items.

Table 6

Spearman Correlation Between Prequiz Self-Efficacy Items and Corresponding AVQUIZ Items

PreQuiz Item	Multiple-Choice	Constructed-Response
1. A rectangular shaped fuel tank measures	.353**	.524**
2. An aircraft as loaded weighs	.173	.313**
3. As weighed, the total empty weight of an aircraft is	.067	.169
4. An aircraft with an empty weight of	.361**	.487**

Descriptive Statistics

Descriptive statistics by response format (multiple-choice vs. constructed-response) and group (word vs. list) are provided in Table 7. Mean scores for the multiple-choice format significantly exceeded those for the constructed-response format. Although mean scores for the word problem group exceeded the group that received abbreviated list problems, differences were not significant.

Reliability Estimates

Coefficient alpha was used to estimate the internal consistency of the dichotomously scored questions on the aviation mechanic AVQUIZ. Cronbach's alpha was satisfactory (α= .73). When analyzed by question context, the traditional word sentence format questions (α = .78) had higher reliability than the abbreviated list format (α = .59).

Table 7

Mean Percent Correct by Response Type and Group

Response Type	Group	M	SD	n
Multiple Choice	Word	54.94	25.00	54
	List	46.38	17.89	46
	Total	51.00	22.33	100
Constructed Response	Word	30.86	29.21	54
	List	23.55	24.99	46
	Total	27.50	27.46	100

Repeated Measures Design

The repeated measures design is appropriate when the same subjects are given a series of tests or subtests (Stevens, 1990). The one between one within design is sometimes called by other names, such as Lindquist Type I, split plot, or two way ANOVA, with repeated measures on one factor (Stevens, 1990). "The benefit of a within-subjects design is each subject serves as his/her own control, reducing extraneous error variance and reducing the total number of subjects required. The benefit of the between factor is each subject is not required to serve in multiple experimental conditions, reducing risks of carryover effects in the study and minimizing exposure to treatments and excessive measurements (Cruz, 1997, p. 1)."

The assumptions for a repeated measures analysis are independence of observations, multivariate normality, and sphericity. The results of the univariate and multivariate analysis methods are equivalent for models with only two levels of the repeated factor (Cruz, 1997). Also when the repeated factor has only two levels, sphericity is always satisfied (Cruz, 1997). Because there were group differences on self-efficacy, that variable was used as a covariate.

These analyses are reported in terms of each of the research questions they address.

Research Question 1

Does response format (multiple-choice versus constructed-response) affect achievement on aviation weight and balance problems, controlling for self-efficacy?

A 2 x 2 repeated measures analysis of covariance (ANCOVA) revealed a significant within-groups difference for the dependent variable of response format (multiple-choice versus constructed-response). The main effect of format was significant, $F(1,99)=32.09$, $p < .01$, $\eta^2 = .25$. As expected, the multiple-choice scores ($M = 51.00$, $SD = 22.32$, $n = 100$) were higher than the constructed-response scores ($M = 27.500$, $SD = 27.46$, $n = 100$).

Research Question 2

Does question context (word story problem versus abbreviated list) affect achievement on aviation weight and balance problems, controlling for self-efficacy?

Question context (word problem versus abbreviated list) did not have a significant effect on achievement, $F(1,99) = 3.17$, $p = .08$, $\eta^2 = .03$. Individuals who received typical word problems ($M = 42.90$, $SD = 24.78$, $n = 54$) scored higher than those who received the abbreviated list format ($\underline{M}$ = 34.96, $\underline{SD}$ = 18.73, $\underline{n}$ = 46), but this difference was not statistically significant.

Aviation mechanics use and read acronyms and abbreviations in their daily tasks. The students were generally comfortable with the abbreviated list problems. This finding indicates that changing question context on complex, computational questions to a more condensed format would probably not create a problem for students on their FAA certification exams. The questions were clear and understandable without elaboration. With the exception of two questions, response times were less for the abbreviated list context. This could be an important advantage on exams with 50 to 100 questions.

Research Question 3

Does the interaction of response format and question context affect achievement on aviation weight and balance problems, controlling for self-efficacy?

The interaction effect between response format and question context, controlling for self-efficacy, was not statistically significant, $F\ (1,99) = .178$, $p = .67$, $\eta^2 < .01$.

(insert Figure 1)

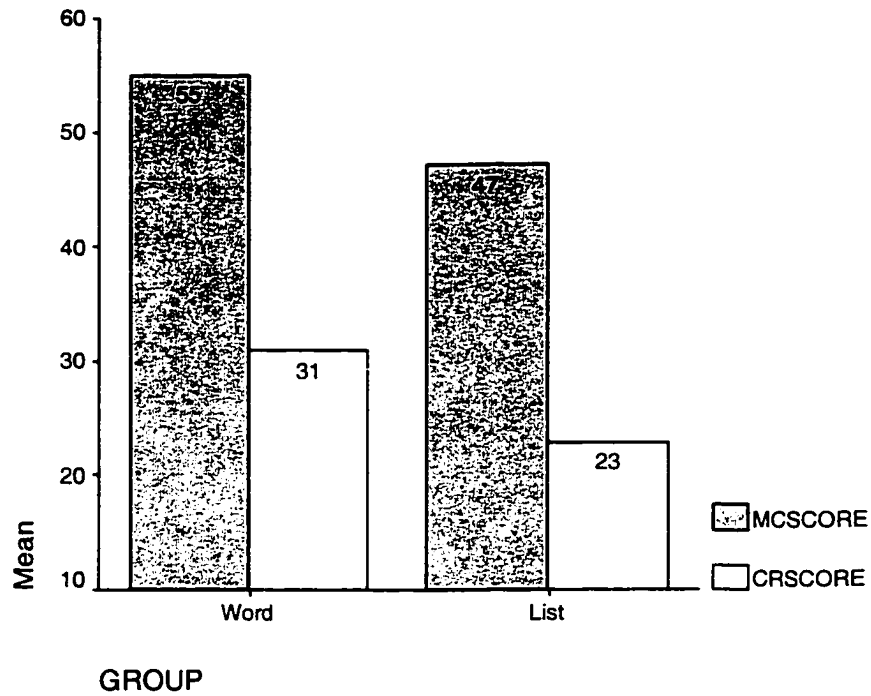


Figure 1. Relationship between response format and question context.

Item Analyses

Tables 8 and 9 list the means and standard deviations for each question. Questions 8 and 11 were extremely difficult as a whole, regardless of question context. Distractor analysis of their matching multiple-choice questions 3 and 6 revealed the answers were evenly divided between the three options. Thus, the multiple-choice questions provided a real challenge for the students.

Question 8 tests knowledge of FAA regulations pertaining to which fluids are considered part of the aircraft's empty weight. Analysis of student answers revealed the following: 7 students correctly solved the problem; 13 students added potable water instead of subtracting it from the empty weight; 12 students subtracted both the potable water and the hydraulic fluid; and an additional 10 students ignored both the

Table 8

Mean Scores of Multiple-Choice Questions

#	Group	M	SD	n
1	Word	.76	.43	54
	List	.65	.48	46
	Total	.71	.46	100
2	Word	.65	.48	54
	List	.61	.49	46
	Total	.63	.49	100
3	Word	.46	.50	54
	List	.41	.50	46
	Total	.44	.50	100
4	Word	.41	.50	54
	List	.28	.46	46
	Total	.35	.48	100
5	Word	.65	.48	54
	List	.59	.50	46
	Total	.62	.47	100
6	Word	.37	.49	54
	List	.28	.46	46
	Total	.32	.47	100

Table 9

Mean Scores of Constructed-Response Questions

#	Group	M	SD	n
7	Word	.67	.48	54
	List	.61	.49	46
	Total	.64	.49	100
8	Word	.11	.32	54
	List	.02	.15	46
	Total	.07	.26	100
9	Word	.28	.45	54
	List	.20	.40	46
	Total	.24	.43	100
10	Word	.35	.48	54
	List	.28	.46	46
	Total	.32	.47	100
11	Word	.11	.32	54
	List	.09	.25	46
	Total	.10	.28	100
12	Word	.33	.48	54
	List	.22	.42	46
	Total	.28	.45	100

potable water and hydraulic fluid to compute the aircraft empty weight. Thus 35 students used an incorrect schema to solve the problem. Four students omitted an answer, and 37 students provided a mathematically implausible answer. An additional 17 students made minor arithmetic errors.

Question 11 is a more complex problem that requires additional computations to compute the aircraft empty weight. Nine students answered correctly; 27 students provided a mathematically plausible answer. Ten students omitted question 11, and 54 students supplied an implausible answer. Seven students subtracted the hydraulic fluid, thus applying an incorrect schema. This subgroup of seven students also applied the same incorrect schema to questions 3, 6, and 8, which are all variations of the same problem. As these students were from the same school, it suggests an instructional error or clear misinterpretation of the FAA regulation.

Response Times

Wyndhamn and Säljö (1997) reported many examples of students who solve problems by following rules and using symbols without reflecting or analyzing the implications of these rules. Evidence of this behavior could be deduced from analysis of response time per question and observation of response patterns. For example, one student answered correctly multiple-choice questions 1, 2 and 5 in 7, 2, and 5 seconds, respectively, which collectively did not provide enough time to read one question in its entirety. Other students repeated the same answer for several constructed-response questions. For example, one student supplied "10" on five consecutive answers; another typed in "98" on each constructed-response answer.

Regardless of test version, students spent more time on the response format they encountered first. Average response times per question are reported in Tables 10 and 11. Question 6 contains 120 words in the stem and all three responses. Approximately 32 people answered question 6 correctly; response time varied between 2 seconds and 780 seconds. Eleven students answered question 6 correctly in less than 10 seconds. The remaining 68 students answered incorrectly with response time varying between 2 seconds and 868 seconds. Another 32 students answered B, which was the answer before the numbers were changed for the quiz. Evidently, the students recognized the problem and former answer from the published item bank.

Question 11 is the constructed-response counterpart to question 6. Question 11 has 114 words in the stem. Nine people answered question 11 correctly, with response time ranging from 192 to 538 seconds. Question 11 is the published question without the three responses. The correct answer "60.31" is distractor B on question 6. Nine people omitted question 11. Answers ranged from 0 to 8946.3.

Six students completed the quiz in less than 5 minutes. Scores ranged from 8 percent to 42 percent. The student with the score of 8 percent answered one of the multiple-choice questions correctly, omitted all six constructed-response questions, omitted all demographic questions, completed no notes, yet rated himself moderately confident on all four prequiz items. The student who scored 42 percent answered five of six multiple-choice questions correctly, and none of the constructed-response items correctly. He supplied the same answer "10" for questions 8 through 12. He also rated himself as "Fair" in English proficiency and "low confidence" on all four prequiz items. Five of the six students who completed the quiz in less than 5 minutes answered all six constructed-response questions incorrectly.

Four students took more than 1 hour to complete the quiz. Their scores ranged from 33 percent to 75 percent. All were older male students (22 and older), and they represented all four schools. The students set up the problems correctly but made arithmetic errors. On question 6, their response times ranged from 413 to 647 seconds. On question 11, their response times ranged from 150 to 2125 seconds.

Three students scored 100 percent on the quiz. Total response times for these individuals varied from 35 to 49 minutes. On question 6, response times ranged from 173 to 515 seconds. On question 11 response times ranged from 230 to 518 seconds. As mentioned earlier, questions 6 and 11 are parallel items. Response format did not appear to influence the outcome or problem-solving strategies used by high scoring individuals.

<u>Student Feedback</u>

During administration of the quiz, the researcher received feedback from students and instructors at the schools. The most common comments were "The test was too hard" or "The test was too long." Some said: "I didn't do very well" or "I got tired." "We studied weight and balance months ago." "I don't remember much of that."

A review of student notes was attempted to identify why so many students performed poorly on the quiz. Examples of student notes can be found in Appendix F. Although many students provided computations of each question, there were a large number of students that provided computations on only one or two questions.

Table 10

Response Times of Multiple-Choice Questions

#	Group	M	SD	n
1	Word	136.59	102.33	54
	List	109.96	85.34	46
	Total	124.34	95.30	100
2	Word	162.00	140.48	54
	List	120.50	114.17	46
	Total	142.91	130.08	100
3	Word	134.56	136.76	54
	List	95.20	97.51	46
	Total	116.45	121.34	100
4	Word	144.02	158.49	54
	List	165.63	218.87	46
	Total	153.96	187.99	100
5	Word	137.98	156.28	54
	List	131.48	150.86	46
	Total	134.99	153.07	100
6	Word	147.43	210.97	54
	List	176.46	203.61	46
	Total	160.78	207.08	100

Table 11

Response Times of Constructed-Response Questions

#	Group	M	SD	n
7	Word	103.44	73.70	54
	List	67.78	35.67	46
	Total	87.04	61.68	100
8	Word	126.52	86.49	54
	List	111.76	77.38	46
	Total	119.73	82.35	100
9	Word	230.96	289.20	54
	List	148.39	152.15	46
	Total	192.99	238.77	100
10	Word	103.44	92.84	54
	List	81.85	91.89	46
	Total	93.51	92.57	100
11	Word	233.30	236.03	54
	List	200.20	330.64	46
	Total	218.07	282.47	100
12	Word	147.41	121.41	54
	List	144.50	149.72	46
	Total	146.07	134.47	100

Combined with the sparse notes, low test scores and response times, it is clear that many students had great difficulty with the AVQUIZ. As a result of this finding, I ran a second repeated measures analysis of covariance on all students who scored at or above the median score. The main effect of response format was significant, $F(1,60) = 25.80$, $p < .01$, $\eta^2 = .31$. The main effect of question context was not significant, $F(1,60) = .78$, $p = .45$, $\eta^2 = .01$. The interaction effect was also not significant, $F(1,60) = 2.52$, $p = 12$, $\eta^2 = .04$. As these results reinforced the original analysis, I reviewed other factors, such as reading comprehension, age, and educational background.

Reading Comprehension

Five students marked that English was not their native language. Their exam scores ranged from 17 percent to 50 percent, although most of them rated themselves as proficient in English (marked Good, Very Good, or Excellent), and moderately confident on the prequiz items. The student who scored 17 percent answered only four questions and omitted the other eight questions. Another non-native English student scored 33 percent. He answered four multiple-choice questions correctly and guessed all six constructed-response questions incorrectly. A problem with reading comprehension could have explained these low test scores.

One student commented that the Visual Basic program did not accept plus "+" or minus "-" symbols. He found this omission confusing as aviation type data sheets typically include the symbols to indicate direction. Direction is a significant element in any weight and balance problem. Since other students were able to enter "+" and minus "-" symbols, this may have been a keyboard error. On the constructed-response questions, the program only accepted entries from the numeric keypad.

Wyndhamn and Säljö (1997) reported that students learn to look for an algorithm in word problems as a means of constructing an answer. The word problem becomes a mathematical exercise and loses its contextual embeddedness (p. 366). This could explain why the constructed-response questions received so many unrealistic answers. The students were applying some mathematical formula without any regard to the practicality of the answer. It is possible that failure to solve the problem resulted from the students' lack of comprehension or inability to situate the problem in context rather than lack of mathematical skills.

Age and Educational Factors

Older students (26-29) tended to work longer on the constructed-response problems than younger students. High school students (15-17) quickly read the questions and determined whether they could solve the problem. Students between 18-21 performed best on both types of questions, multiple-choice and constructed-response.

Students with 1-2 years of college scored higher than all other groups on constructed-response questions. A one-way analysis of variance revealed significant differences between groups on both multiple-choice scores [$F(1,6) = 2.61$, $p = .03$] and constructed-response scores [$F(1,6) = 3.75$, $p < .01$]. Post hoc tests indicated significant differences between high school graduates and students with 1-2 years of college. Tables 12 and 13 list the mean scores by age group and educational level.

Table 12

Mean Scores by Age Group

	Multiple-Choice			Constructed-Response		
Age	M	SD	n	M	SD	n
15-17	44.44	22.77	6	22.22	25.09	6
18-21	56.57	26.98	33	36.36	33.97	33
22-25	52.00	21.69	25	30.00	25.46	25
26-29	44.44	20.52	12	29.17	22.61	12
30+	47.72	15.68	22	17.42	18.88	22
N/A[f]	25.00	22.57	2	16.67	23.57	2

[f]N/A = not available

Table 13

Mean Scores by Educational Level

	Multiple-Choice			Constructed-Response		
Educational Level	M	SD	n	M	SD	n
High school student	48.15	22.74	9	18.52	15.47	9
High school graduate	43.33	20.41	25	15.33	24.02	25
GED equivalent	33.33	13.61	4	16.67	23.57	4
1-2 yrs college	56.48	22.53	54	37.65	28.81	54
College graduate	58.33	17.48	6	16.67	10.54	6
N/A[g]			2			2

[g]N/A = not available

Summary

Results from this study indicate response format is a factor in performance on aviation mechanic weight and balance exam questions. The next chapter will discuss the implications of this finding, limitations of the study, and potential areas of future study.

Chapter 5

Discussion

The present study investigated the effects of question context and response format on achievement in an aviation mechanic weight and balance quiz. An alternative question format was presented to simulate a more realistic problem situation. An aviation mechanic quiz was designed to measure the differences in test performance on the subject of weight and balance. Typical quiz items required the student to compute empty weight center of gravity or changes in center of gravity after alterations to existing equipment had been made. Other quiz items required additional knowledge of FAA rules regarding weight and balance measurements or knowledge of Newton's third law of motion.

It was hypothesized that mechanics would perform better on multiple-choice questions because of their familiarity with the FAA airman knowledge test program. It was suggested that self-efficacy could be a source of variance in student test performance. The findings from this study support previous research (Katz, Bennett, and Berger, 2001; Bridgeman, 1992) that the multiple-choice format may be easier than the constructed-response format in solving mathematical problems.

It was hypothesized that aviation mechanics would perform better on questions that resembled real-world problems as opposed to textbook problems. The results did not support the hypothesis that an alternative question context would be easier for aviation mechanics because of its resemblance to real-world problems. To aid in the review of the major findings of this study, I will present a brief summary of the findings regarding each research question.

Summary

Research question 1 -- **Does response format (multiple-choice versus constructed-response) affect achievement on aviation weight and balance problems, controlling for self-efficacy?** There was a significant within-groups difference for the dependent variable of response format (multiple-choice versus constructed-response). The multiple-choice scores were higher than the constructed-response scores. In this study students frequently provided an implausible answer with the constructed-response format. There were no significant differences in response times between formats. This finding suggests that students were not selecting an answer through the process of elimination of the other alternatives.

Analysis of student notes confirmed that many students were performing the computations to solve multiple-choice problems as well as constructed-response problems. Unlike Birenbaum and Tatsuoka's (1987) study, it was easier to diagnose problems with the multiple-choice format because the distractors were based on common misconceptions. It was difficult to establish the problem solving model for many answers supplied to the constructed-response questions. Results from this study were similar to Bridgeman's (1992) study in that questions that were relatively easy to solve in the multiple-choice format became very difficult when transformed into the constructed-response format.

Research question 2 -- **Does question context (word story problem versus abbreviated list) affect achievement on aviation weight and balance problems, controlling for self-efficacy?** In this study, question context did not present a major problem for students. There was no difference in scores between the word problem group and the abbreviated list group. This finding is similar to prior research (Davis-Dorsey, et al. 1991; Moyer et al., 1983) where a simplified language format did not help solve mathematical problems for older and more experienced students. Examination of student notes revealed that students used the same problem representation to solve each problem, regardless of format or question context. Most students used a table structure to represent the weight and balance problems.

Rewording the problems to remove extraneous information did not alter the problem-solving model required. Semantic structure did not appear to be a factor in solving weight and balance problems. There were several examples of simple drawings from low scoring students with no other detailed tables or equations, thus supporting the results from Threadgill-Sowder et al. (1985). Students used the same strategies to solve multiple-choice and constructed-response questions regardless of test version. There was no significant difference in score or response time between the two question contexts.

Research question 3 -- **Does the interaction of response format and question context affect achievement on aviation weight and balance problems, controlling for self-efficacy?** There was no interaction between response format and question context. Thus, only the response format main effect resulted in a significant difference between groups.

The design elements incorporated into this study, namely questions counterbalanced, random assignment to groups, sample size and selection, and covariate controlling for bias, contributed to reducing both internal and external threats to validity.

Possible Explanations for Low Test Performance

Although the findings of this study provided clear answers to the research questions, I was surprised by the low scores on the AVQUIZ. The following observations were made after analyzing student notes, and test and item analyses. Based on these observations, I have generated some plausible explanations for the test difficulty.

All results suggest that the multiple-choice format was easier than the constructed-response format. Students in general were unable to solve constructed-response problems. They were unable to guess the answer through elimination of responses. An advantage of the multiple-choice format was that students could make minor computational mistakes and still select from the closest response. In the constructed-response format, any type of computational error led to a wrong answer. The scoring restriction of exact answers only scored as correct may have led to an underestimation of the constructed-response format.

These findings also suggest that students may not be learning procedures for solving real weight and balance problems, but are often only recalling answers to questions they know will be tested on the FAA certification exam. Several students made no attempt to answer constructed-response questions, even though they had just answered a similar multiple-choice question. College students performed better than high school students or high school graduates on the constructed-choice format. Their training may have included more emphasis on solving real problems than the other groups, and thus were more comfortable with the constructed-response format. Nearly all of the college students had more math and science background and prior civilian aviation experience.

Math computational skills were often weak. Many students set up the problem according to the mathematical model, but obtained the wrong answer through simple arithmetical errors. Combining response formats on the same quiz may have assisted some students. With the multiple-choice questions, they were able to check their work and overcome minor arithmetical errors.

Students frequently failed to recall pertinent FAA regulations. For example, the students added fluids to the aircraft total weight when they should have subtracted or ignored the fluid's weight. Perhaps the students did not have well-developed production systems to enable them to solve the problems. Several students from the same school chose the same incorrect schema to solve similar problems. During the pilot study discussed in Appendix C, I noticed that many students used a table structure with the mnemonic device "W A M" (Weight x Arm = Movement) to solve the weight and balance problems. But the students were inserting all numerics from the question, rather than applying appropriate production rules as discussed earlier in the ACT* theory. They were not testing each condition before performing the action. While the table structure was helpful in representing the problem, it did not substitute for asking relevant "IF... -THEN..." questions about what information should be included in the problem solution.

Some other possible explanations for the poor test results include the following. Too much time elapsed between instruction and testing. Students could not remember how to solve the problems, nor could they remember the formulas needed to solve the problems. The date of testing at one school was the last scheduled day of training, and many students were distracted. At another school, the students had completed the weight and balance section several weeks prior to the quiz. As the quiz did not count as a grade, students were not motivated.

Some students appeared overconfident and failed to recognize the complexity of the problems. They rated themselves as highly confident on the prequiz items yet failed to score above 50 percent on the quiz.

The questions were lengthy and took too much time to solve. When some students finished quickly, other students rushed through the remaining questions to join their peers on break or in the cafeteria. Fatigue seemed to be a factor for a few students. They complained that the test was too long.

Limitations of Study

This sample was limited to four Part 147 schools in Oklahoma and Texas. Three of these schools are career tech programs, and the fourth school is a private university. The career tech schools included both high school and adult students enrolled in daytime and evening classes. The quiz was conducted near or at the end of the school curriculum and may not have immediately followed the instruction over the weight and balance material. Motivation was low at the career tech schools. Students did not solve many problems using the notepaper supplied. Response times were shorter compared to the college sample. Some used onscreen calculators, which may have contributed another source of error because the students did not check their work. The computer program did not allow students to revisit a question once an answer had been selected.

The university selection included all adult students enrolled in a 4-year curriculum that prepared them for a future in the aviation industry. Several had pilot licenses and considerable aviation experience. The quiz followed immediately a block of instruction on weight and balance. The students knew prior to the quiz that it consisted of multiple-choice and constructed-response questions. Motivation and self-efficacy was very high for this group.

Although Part 147 schools all share the same basic curriculum, these four schools may not represent the full range of classroom/laboratory instruction that is given to aviation mechanics. Each school provides its own tools, equipment, and airplane parts for student use. The four schools represented a cross-section of Part 147 schools, with adequate access to excess airplanes and airplane parts from military and private industry.

Data collection was restricted to computer-generated data saved onto a floppy disk. Three of the four schools had sophisticated instructional technology (IT) programs. Their computer labs were new and used the Windows 2000 operating system. The fourth school had four computers located in the aviation technology hangar and used Windows 95. Modifications were made to the Visual Basic program after the developmental tryout and pilot study; however, the program failed to collect some data during the first administration of the study. Re-ordering multiple-choice and constructed-response questions caused additional problems in this first administration.

Failing to provide necessary formulas in ballast questions contributed to the low scores on two questions. These two questions provided the greatest source of difficulty on the quiz. Instructors and subject matter experts commented that they would not expect someone to have memorized the formula to answer these questions.

The Visual Basic program did not collect response time on the two practice items or from the prequiz items. From the student notes, many students devoted considerable time and effort to draw triangles and trapezoids, which were primarily mathematics problems, and not weight and balance problems. The prolonged time spent on the prequiz and practice items could have contributed to some of the fatigue students mentioned after the quiz. Their notes almost always contained the practice questions, with the rest of the quiz slowly tapering off. Response times decreased toward the end of the quiz.

The results of this study may not generalize to other subject areas, such as basic electricity and aircraft systems, or to different populations, such as student pilots. Mechanics receive more instructional hours on basic electricity and aircraft systems. Problems in basic electricity require different knowledge and skills. Basic electricity computations generally require only one mathematical computation, as opposed to the multi-step weight and balance problems. Pilots are directly affected by weight and balance measurements and may receive more instruction in this subject. Pilots may also be more comfortable with weight and balance computations.

The study did not collect reading comprehension or math achievement scores from another source to identify which individuals would experience difficulty with the quiz. In this study some students were initially confused by the prequiz items. They tried to solve the problems instead of rating their confidence. Verbal and written instructions, in addition to the computer example, should be incorporated into future administrations.

Implications for FAA Practice

Changing the FAA certification exam to include constructed-response format for computation questions would be impractical because of the high volume of exams that are delivered annually. The test scoring process requires that tests are scored immediately and that the applicant receive a printed test report before leaving the test center. The scannable format described in Bridgeman's (1992) study would not provide immediate results and would require additional resources. However, the FAA should change the computation questions on exams, publishing only sample questions on the Internet. All formal training courses would be compelled to teach procedures rather than teaching the test.

Training should include more review and practice throughout the Part 147 school curriculum. Systematic review of prior blocks of instruction should be incorporated routinely. For the most part, students are not exposed to question formats that require them to supply an answer. Students are taught general strategies and algorithms to solve weight and balance problems. They are encouraged to estimate answers. This policy must change.

Aviation mechanic students know that multiple-choice word problems are used on all FAA certification exams. FAA question banks are available through the Internet. Test questions are also available in commercially published textbooks that provide expert opinions regarding the correct answers. Included in most Part 147 school curricula are sample computer tests composed of actual FAA test questions.

The FAA exams must also change so that students do not see actual test problems and memorize answers before being tested.

The students in this study used the same strategies on all the weight and balance problems, regardless of the school they attended. The same schema "Weight x Arm = Moment" occurred repeatedly in their notes. Students used the same table structure to perform the arithmetic computations. They should be taught to question each fact presented in the problem. Real-world problems are rarely straightforward. The majority of errors students made resulted from computation mistakes or from misinterpreting the FAA regulation. These types of errors can easily be corrected.

Implications of Future Research

Although students were confident that they could solve typical weight and balance problems, mean scores of both multiple-choice and constructed-response questions were disappointingly low. A follow-up study should be conducted with students enrolled in Part 141 pilot schools. Weight and balance is a critical subject area for both pilots and mechanics; thus it is important that instructors teach procedural knowledge and that test writers generate challenging, realistic problems to ensure aviation professionals have the necessary skills to do their job.

To improve this study, I recommend the following changes to the test instrument. Practice questions should mirror basic weight and balance questions on the quiz. Ballast questions should be revised to include necessary formulas. The quiz should be reduced to no more than 8 to 10 questions. A standardized measure of reading comprehension should be obtained prior to the study in order to identify another source of preexisting bias.

Training differences between career tech schools and colleges should be further investigated. More information should be obtained regarding the weight and balance curriculum at each participating school. Are constructed-response questions routinely included in the classwork assignment? Are remedial math and reading courses offered to older students?

Summary

The general research question guiding my research was: Is there a difference in cognitive test performance when participants use either multiple-choice or constructed-response questions, or when the item type is a story problem or an abbreviated list? The results clearly showed that response format was a significant factor and that question context did not affect test performance for aviation mechanic students in solving weight and balance problems. Students solved the abbreviated list items used on this quiz about as well as they did the word problems. It appears that either format could be used in future testing with few difficulties. However, students performed poorly on the constructed-response format, which may explain why the test scores were lower than expected. Additional research is needed to determine whether the inability to answer constructed response items is a serious problem or merely a result of the experimental conditions in this study.

References

Ackerman, T. A., & Smith, P. L. (1988). A comparison of the information provided by essay, multiple-choice, and constructed-response writing tests. Applied Psychological Measurement, 12 (2), 117-128.

Aiken, L. R. (1982). Writing multiple-choice items to measure higher-order educational objectives. Educational and Psychological Measurement, 42, 802-806.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1992). Automaticity and the ACT* theory. American Journal of Psychology, 105 (2), 165-180.

Anderson, J. R. (1995). Cognitive psychology and its implications. New York: W. H. Freeman and Co.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. Psychological Review, 84 (2), 191-215.

Barnett-Foster, D. (1993). A comparison of undergraduate test response strategies for multiple choice and constructed response test questions. (Doctoral Dissertation, University of Toronto, 1993). Dissertation Abstracts International, 55, (03A), 0476.

Bennett, R. E. (1993). On the meaning of constructed response. In R. E. Bennett & W. C. Ward (Eds.), Construction versus choice in cognitive measurement (pp. 1-22). Hillsdale, NJ: Erlbaum.

Bennett, R. E., Rock, D. A., & Wang, M. (1991). Equivalence of free-response and multiple-choice items. Journal of Educational Measurement, 28 (1), 77-92.

Birenbaum, M., & Tatsuoka, K. (1987). Open-ended versus multiple-choice response formats – it does make a difference. Applied Psychological Measurement, 11 (4), 385-395.

Boodoo, G. (1993). Performance assessments or multiple choice. Educational Horizons, 72, 50-56.

Braswell, J., & Kupin, J. (1993). Item formats for assessment in mathematics. In R. E. Bennett & W. C. Ward (Eds.), Construction versus choice in cognitive measurement (pp. 167-182). Hillsdale, NJ: Erlbaum.

Bridgeman, B. (1992). A comparison of quantitative questions in open-ended and multiple-choice formats. Journal of Educational Measurement, 29, 253-271.

Campbell, J. R. (1999). Cognitive processes elicited by multiple-choice and constructed-response questions on an assessment of reading comprehension. (Doctoral Dissertation, Temple University, 1999). Dissertation Abstracts International, 60, (7A), 2428.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), Advances in the psychology of human intelligence (pp. 17-76). Hillsdale, NJ: Erlbaum.

Choi, J., & Hannafin, M. (1997). The effects of instructional context and reasoning complexity on mathematics problem-solving. Educational Technology Research and Development, 45 (3), 43-55.

Collis, K., & Romberg, T. A. (1991). Assessment of mathematical performance: an analysis of open-ended test items. In M. C. Wittrock & E. L. Baker (Eds.), Testing and cognition (pp. 82-130). Englewood Cliffs, NJ: Prentice-Hall.

Cruz, C. E. (1997). A comparison of univariate and multivariate analysis of variance for human performance data in a repeated measures design. Unpublished doctoral dissertation, University of Oklahoma, Norman.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R.(1988). The role of understanding in solving word problems. Cognitive Psychology, 20, 405-438.

Davis-Dorsey, J., Ross, S. M., & Morrison, G. R. (1991). The role of rewording and context personalization in the solving of mathematical word problems. Journal of Educational Psychology, 83 (1), 61-68.

DeCorte, E., Verschaffel, L., & De Win, L.(1985). Influence of rewording verbal problems on children's problem representations and solutions. Journal of Educational Psychology, 77 (4), 460-470.

Ebel, R., & Frisbie, D. A. (1991). Essentials of educational measurement (5th Ed.). Englewood Cliffs, NJ: Prentice-Hall.

Federal Aviation Administration. (1999). Aircraft weight and balance handbook, FAA-H-8083-1. Pittsburgh, PA: U. S. Government Printing Office.

Federal Aviation Administration. (1997). Pilot's handbook of aeronautical knowledge, AC-61-23C (Rev. ed.). Pittsburgh, PA: U. S. Government Printing Office.

Frederiksen, N. (1984). The real test bias: influences of testing on teaching and learning. American Psychologist, 39, 193-202.

Gagne, E. D. (1993). The cognitive psychology of school learning. New York: Harper Collins.

Greeno, J. G. (1978). Understanding and procedural knowledge in mathematics instruction. Educational Psychologist, 12, 262-283.

Hackett, G., Betz, N. E., O'Halloran, M. S., & Romac, D. S. (1990). Effects of verbal and mathematics task performance on task and career self-efficacy and interest. Journal of Counseling Psychology, 37 (2), 169-177.

Haladyna, T. A. (1997). Writing test items to evaluate higher order thinking. Hillsboro, NJ: Erlbaum.

Hinsley, D. A., Hayes, J. R., & Simon, H. A. (1977). From words to equations: Meaning and representation in algebra word problems. In P. Carpenter & M. Just (Eds.), Cognitive Processes in Comprehension (pp. 89-105). Hillsdale, NJ: Erlbaum.

Katz, I. R., Bennett, R. E., & Berger, A. E. (2000). Effects of response format on difficulty of SAT-mathematics items: It's not the strategy. Journal of Educational Measurement, 37(1), 39-57.

Kintsch, W. (1998). Comprehension: a paradigm for cognition. New York: Cambridge University Press.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. Psychological Review, 92 (1), 108-129.

Kintsch,W., & van Dijk, T.A.(1978). Toward a model of text comprehension and production. Psychological Review, 85 (5), 363-394.

Larkin, J. H. (1983). The role of problem representation in physics. In D. Genter & A. L. Stevens (Eds.), Mental models (pp. 75-98). Hillsdale, NJ: Erlbaum.

Latterel, C. M. (2000). Assessing NCTM standards-oriented and traditional students' problem solving ability using multiple-choice and open-ended questions. (Doctoral Dissertation, University of Iowa, 2000). Dissertation Abstracts International, 61, (8-A), 3095.

Luekele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed-response, and examinee-selected items on two achievement tests. Journal of Educational Measurement, 31 (3), 234-250.

Martinez, M. E. (1999). Cognition and the question of test item format. Educational Psychologist, 34 (4), 207-218.

Mayer, R. E. (1978). Qualitatively different storage and processing strategies used for linear reasoning tasks due to meaningfulness of premises. Journal of Experimental Psychology, 4, 5-18.

Mayer, R. E. (1982). Different problem-solving strategies for algebra word and equation problems. Journal of Experimental Psychology, 8 (5), 448-462.

Mayer, R.E., Larkin, J.H., & Kadane, J.B. (1984). A cognitive analysis of mathematical problem-solving ability. In R. J. Sternberg (Eds.), Advances in the psychology of human intelligence (pp. 231-273). Hillsdale, NJ: Erlbaum.

Messick, S. (1993). Trait equivalence as construct validity of score interpretation across multiple methods of measurement. In R. E. Bennett & W. C. Ward (Eds.), Construction versus choice in cognitive measurement (pp. 61-73). Hillsdale, NJ: Erlbaum.

Moyer, J. C., Sowder, L., Threadgill-Sowder, J., & Moyer, M. B. (1983). Story problem formats: drawn vs. verbal vs. telegraphic. Journal for Research in Mathematics Education, 15 (5), 342-351.

Pajares, F., & Miller, M. D. (1997). Mathematics self-efficacy and mathematical problem solving implications of using different forms of assessment. Journal of Experimental Education, 65 (3), 213-228.

Reusser, K. (1990). Textual and situational factors in solving mathematical word problems. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Riley, M. S., Greeno, J. G., & Hinsley, D. A. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), The development of mathematical thinking (pp. 153-196). New York: Academic Press.

Ryan, A. M., & Greguras, G. J. (1998). Life is not multiple choice: reactions to the alternative. In M. Hakel (Ed.), Beyond multiple choice: evaluating alternatives for traditional testing for selection (pp. 183-202). Mahwah, NJ: Erlbaum.

Sebrechts, M. M., Enright, M., Bennett, R. E., & Martin, K. (1996). Using algebra word problems to assess quantitative ability: attributes, strategies, and errors. Cognition and Instruction, 14 (3), 285-343.

Stevens, J. P. (1990). Intermediate statistics. A modern approach. Hillsdale, NJ: Erlbaum.

Snow. R. E. (1993). Construct validity and constructed-response tests. In R. E. Bennett & W. C. Ward (Eds.), Construction versus choice in cognitive measurement (pp. 46-59). Hillsdale, NJ: Erlbaum.

Standards for educational and psychological testing. (1985). Washington, DC: American Psychological Association.

Thorndike, R. L. (1971). Educational measurement (2nd ed.) Washington, D. C. American Council on Education.

Threadgill-Sowder, J., Sowder, L., Moyer, J. C., & Moyer, M. B. (1985). Cognitive variables and performance on mathematical story problems. Journal of Experimental Education, 54, 56-62.

Traub, R. (1993). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. In R. E. Bennett & W. C. Ward (Eds.), Construction versus choice in cognitive measurement (pp. 29-43). Hillsdale, NJ: Erlbaum.

Traub, R., & Fisher, C. W. (1977). On the equivalence of constructed-response and multiple-choice tests. Applied Psychological Measurement, 1 (3), 355-369.

Van Lehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), Foundations of cognitive science (pp. 527-579). Cambridge, MA: MIT Press.

Ward, W. C. (1982). A comparison of free-response and multiple-choice forms of verbal aptitude tests. Applied Psychological Measurement, 6 (1), 1-11.

Ward, W. C., Frederiksen, N., & Carlson, S. B. (1980). Construct validity of free-response and multiple-choice versions of a test. Journal of Educational Measurement, 17, 11-29.

Wyndhamn, J., & Saljo, R. (1997). Word problems and mathematical reasoning – a study of children's mastery of reference and meaning in textual realities. Learning and Instruction, 7 (4), 361-382.

Appendix A

Weight and Balance

The following pages present an oversimplified explanation of weight and balance theory as it applies to aviation mechanics. Weight and balance computations involve simple mathematical concepts and operations and procedural knowledge of FAA regulations. Information has been excerpted from the following FAA publications: Pilot's handbook of aeronautical knowledge, AC 61-23C, 1997, and Aircraft weight and balance handbook, FAA-H-8083-1, 1999.

To understand how weight and balance affect an airplane in flight, the examinee must be able to conceptualize Newton's third law of motion. This physics principle states that for every action there is an equal and opposite reaction. When in flight, there are certain forces acting on the airplane. Among the aerodynamic forces acting on the airplane, lift and gravity (weight) are opposing forces. Lift is the upward force created by the airplane when moved through the air. Gravity (weight) is the downward force that tends to draw the airplane vertically toward the center of the earth. The airplane's center of gravity (CG) is the point on the airplane at which all weight is concentrated (Pilot's handbook of aeronautical knowledge, AC 61-23C, 1997, p. 1-4).

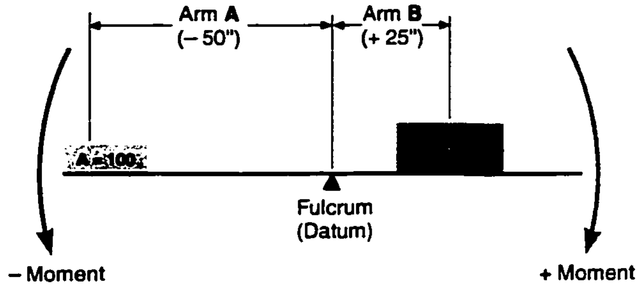
Balance refers to the location of the center of gravity of an airplane. The center of gravity is a point at which an airplane would balance if it were suspended at that point. The primary concern of airplane balance is the fore and aft locations of the center of gravity along the longitudinal axis (imaginary line from the nose to the tail). The location of the center of gravity depends upon the location and weight of the load placed in the airplane. The exact location of the center of gravity is important during flight, because of its effect on airplane stability and performance. As variable load questions are shifted, if the center of gravity of an airplane is displaced too far forward or aft of the longitudinal axis, an unstable condition could result in which the pilot could not control the airplane.

All weight and balance problems are based on the physical law of the lever. Simply stated, "a lever is balanced when the weight on one side of the fulcrum multiplied by its arm is equal to the weight on the opposite side multiplied by its arm" (Aircraft weight and balance handbook, 1999, p. 2-2). The lever is balanced when the algebraic sum of the moments about the fulcrum is zero. The balanced condition results when the positive moments (those that try to rotate the lever clockwise) are equal to the negative moments (those that try to rotate the lever counterclockwise). See Figure A1.

Basic Principles of Weight and Balance Computations

Total weight can be determined by adding the weight of the empty airplane and everything loaded onto the airplane. The weight must be distributed and balanced around the center of gravity, which is the imaginary point where all the weight is concentrated. The center of gravity range is a safe zone within which the center of gravity must fall; its extremities are called the forward CG limit and aft CG limit. These limits are specified in inches along the longitudinal axis of the airplane, measured from the reference datum. The mechanic who maintains the aircraft and performs the maintenance inspections keeps the weight and balance records current, recording any changes that have been made because of repairs or alterations.

(insert Figure A1)



<u>Figure A1.</u> The lever is balanced when the algebraic sum of the moments is zero.

Appendix B

Mathematical Problem Solving Model

Table B1

Question1 (Multiple-Choice Format)

How many gallons of fuel will be contained in a rectangular-shaped tank which measures two feet in width, three feet in length, and one foot eight inches in depth? (Note: Seven and one-half gallons = one cubic foot)

- ° 66.6 gallons
- ° 75 gallons
- ° 45 gallons

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic measurement terms: length, width, depth, volume, gallons, inches, cubic feet
Understanding	Schematic	Volume of rectangle = length x width x depth
Planning	Strategic	Solve for volume, then multiply by number of gallons in 1 cubic feet
Execution	Algorithmic	2 x 3 x 20/12 x 7.5 = 75

Note. Selecting incorrect schema, i.e., multiplying length by width only, will result in answer C. Converting depth measurement incorrectly to feet could result in answer A.

Table B2

Question 2 (Multiple-Choice Format)

Two boxes that weigh 10 pounds and 15 pounds are placed in an airplane so that their distance aft from the center of gravity are 7 feet and 2 feet respectively. A 5-pound box is placed forward of the center of gravity 3 feet. How far forward of the center of gravity should a fourth box, weighing 20 pounds, be placed so that the center of gravity will not be changed?

- o 3.5 feet
- o 4.25 feet
- o 4.5 feet

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic weight & balance terms: aft, forward, center of gravity
Understanding	Schematic	Newton's third law of motion
Planning	Strategic	Solve for location of box 4 by drawing figure on continuum.
Execution	Algorithmic	$20X + (5 \times 3) = (10 \times 7) + (15 \times 2)$

Note. Division error will result in answer C.

Table B3

<u>Question 3 (Multiple-Choice Format)</u>

As weighed, the total empty weight of an aircraft is 5,862 pounds with a moment of 885,957 pound-inches. However, when the aircraft was weighed, 20 pounds of potable water were on board at 84 inches, and 23 pounds of hydraulic fluid were in a tank located at 101 inches. What is the empty weight center of gravity of the aircraft?

- 150.700 inches
- 151.700 inches
- 151.365 inches

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic weight and balance terms: total empty weight, moment, potable water, hydraulic fluid, empty weight center of gravity
Understanding	Schematic	Weight x Arm = Moment
Planning	Strategic	Set up table structure and solve for missing items.
Execution	Algorithmic	Perform computations; add/subtract appropriate components.

<u>Note</u>. Incorrect schema results in answers A and B. Adding potable water instead of subtracting water will result in answer A; adding water and hydraulic fluid will result in answer B.

Table B4

Question 4 (Multiple-Choice Format)

An aircraft as loaded weighs 2,500 pounds at a center of gravity of 38.4 inches. The center of gravity range of the aircraft is 40.9 to 46.0 inches. Find the minimum weight of the ballast necessary to bring the center of gravity within the given range. The ballast arm is located at 120 inches.

- 63.21 pounds
- 69.53 pounds
- 79.01 pounds

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic terms: center of gravity, center of gravity range, ballast, arm
Understanding	Schematic	$\frac{\text{Empty weight x Distance out of range}}{\text{Distance ballast to new CG}}$
Planning	Strategic	Set up equation.
Execution	Algorithmic	Perform computations: $\frac{2500 \times 2.5}{120 - 40.9}$

Note. Transcription error in multiplication of numerator will result in answer A.

Table B5

Question 5 (Multiple-Choice Format)

An aircraft had an empty weight of 2,886 pounds with a moment of 101,673.78 before several alterations were made. The alterations included: removing two passenger seats weighing 15 pounds each at 71 inches; installing a cabinet weighing 97 pounds at 71 inches; installing a seat and safety belt weighing 20 pounds at 71 inches, and installing radio equipment weighing 30 pounds at 94 inches. What is the new empty weight center of gravity?

- 33.20
- 36.85
- 37.26

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic terms: empty weight, moment, empty weight center of gravity
Understanding	Schematic	Weight x Arm = Moment
Planning	Strategic	Set up problem in table format.
Execution	Algorithmic	Perform computations; add/subtract appropriate components.

Note. Error in reading comprehension (subtracting weight of one seat only) will result in answer C.

Table B6

<u>Question 6 (Multiple-Choice Format)</u>

Given an aircraft where the datum is forward of the main gear center point at 30.15 inches, and the actual distance between tail gear and main gear center points is 360.65 inches. Also the net weight at the right main gear is 9,750 pounds, while the net weight at the left main gear is 9,960 pounds, and the net weight at the tail gear is 1,940 pounds. These items were in the aircraft when weighed: full lavatory water tank weighing 32 pounds at 325 inches, hydraulic fluid weighing 22 pounds at minus eight inches, and removable ballast weighing 150 pounds at 380 inches. What is the empty weight center of gravity of the aircraft?

- 59.86 inches
- 60.31 inches
- 62.44 inches

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic terms: datum, forward, main/left/right/tail gears, hydraulic fluid, lavatory water tank, minus, removable ballast
Understanding	Schematic	Weight x Arm = Moment
Planning	Strategic	Set up problem in table format.
Execution	Algorithmic	Perform computations; add/subtract appropriate components.

<u>Note</u>. Applying wrong rule will result in answers B or C. Removing ballast will result in B; answer C is EWCG before removal of items. Reading comprehension error will result by not correctly interpreting how to compute the tail gear arm.

Table B7

Question 7 (Constructed-Response Format)

A rectangular-shaped fuel tank measures 37 and one-half inches in length, 14 inches in width, and eight and one-fourth inches in depth. How many cubic inches are within the tank?

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic measurement terms: length, width, depth, inches, cubic inches
Understanding	Schematic	Volume of rectangle = length x width x depth
Planning	Strategic	Solve for volume.
Execution	Algorithmic	37 ½ x 14 x 8 ¼

Table B8

<u>Question 8 (Constructed-Response Format)</u>

The total empty weight of an aircraft is 1,800 pounds with a moment of 56,700 pound-inches. When the aircraft was weighed, 18 pounds of potable water were on board at 21.5 inches, and 12 pounds of hydraulic fluid were in a tank located at 24 inches. What is the empty weight center of gravity of the aircraft?

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic weight and balance terms: total empty weight, moment, potable water, hydraulic fluid, empty weight center of gravity
Understanding	Schematic	Weight x Arm = Moment
Planning	Strategic	Set up table and solve for missing items.
Execution	Algorithmic	Perform computations; subtract potable water; leave hydraulic fluid.

Table B9

Question 9 (Constructed-Response Format)

An aircraft as loaded weighs 4,954 pounds at a center of gravity of 30.5 inches. The center of gravity range of the aircraft is 32.0 inches to 42.1 inches. Find the minimum weight of the ballast necessary to bring the center of gravity within the given range. The ballast arm is located at 162 inches.

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic terms: center of gravity, center of gravity range, ballast, arm
Understanding	Schematic	$\frac{\text{Empty weight x Distance out of range}}{\text{Distance ballast to new CG}}$
Planning	Strategic	Set up table and solve for missing items.
Execution	Algorithmic	Perform computations.

Table B10

Question 10 (Constructed-Response Format)

Two boxes which weigh 10 pounds and five pounds are placed in an airplane so that their distance from the center of gravity are 4 feet and 2 feet respectively. How far forward of the center of gravity should the third box, weighing 20 pounds, be placed so that the center of gravity will not be changed?

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic weight and balance terms: forward, center of gravity
Understanding	Schematic	Newton's third law of motion
Planning	Strategic	Solve for location of box 3 by drawing figure on continuum.
Execution	Algorithmic	20X = (10 x 4) + (5 x 2)

Table B11

<u>Question 11 (Constructed-Response Format)</u>

Given an aircraft where the datum is forward of the main gear center point at 30.24 inches, and the actual distance between tail gear and main gear center points is 360.26 inches. Also the net weight at the right main gear is 9,980 pounds, while the net weight at the left main gear is 9,770 pounds, and the net weight at the tail gear is 1,970 pounds. These items were in the aircraft when weighed: full lavatory water tank weighing 34 pounds at 352 inches, hydraulic fluid weighing 22 pounds at minus eight inches, and removable ballast weighing 146 pounds at 380 inches. What is the empty weight center of gravity of the aircraft?

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic terms: datum, forward, main/left/right/tail gears, hydraulic fluid, lavatory water tank, minus, removable ballast
Understanding	Schematic	Weight x Arm = Moment
Planning	Strategic	Set up problem in table structure.
Execution	Algorithmic	Perform computations; add/subtract/omit appropriate components.

Table B12

<u>Question 12 (Constructed-Response Format)</u>

An aircraft with an empty weight of 2,100 pounds and an empty weight center of gravity of 32.5 inches was altered as follows: two 18-pound passenger seats located at 73 inches were removed; structural modifications were made at 77 inches increasing weight by 17 pounds; a seat and safety belt weighing 25 pounds were installed at 74.5 inches; and radio equipment weighing 35 pounds was installed at 95 inches. What is the new empty weight center of gravity?

Phase	Type of Knowledge	Example from Problem
Translation	Linguistic & factual	Basic terms: empty weight, moment, empty weight center of gravity
Understanding	Schematic	Weight x Arm = Moment
Planning	Strategic	Set up problem in table format.
Execution	Algorithmic	Perform computations; add/subtract appropriate components.

Appendix C

Pilot Study

A pilot study was conducted in January 2001 to determine whether question context or response format affected achievement scores on a mathematical problem-solving quiz. The quiz was designed to measure the knowledge and skills that aviation mechanic students use to solve weight and balance problems.

Participants

Twenty-five high school students enrolled in the aviation mechanic curriculum at a midsize community college located in Tulsa, Oklahoma, participated in the pilot study. Three females and 22 males from five separate classes rotated through the computer laboratory. Eighteen students listed age as 15 to 17, and the remaining eight students listed the 18 to 21 age group. Approximately 22 listed algebra, 18 listed geometry, 4 listed trigonometry, and 6 listed physics for prior math and science selections.

Procedure

Because many of the students were under 18, they were briefed prior to the study by their instructors and asked to submit signed parental consent forms, in addition to individual consent forms on the date of the pilot study. The quiz had been postponed twice due to inclement weather in the metropolitan area. The study was conducted on the final day of the semester before final exams.

The initial demographics were revised in the Visual Basic application to include an additional age group (15 – 17), and an additional education level (high school student). The other demographic questions were not changed. In addition to verbal instructions at the beginning of the quiz, the researcher provided a written instruction sheet to each student to clarify how to begin the quiz. The instruction sheet minimized the confusion that students experienced during the developmental tryout.

Students completed one of two versions of a specially designed experimental quiz – AVQUIZ. The quiz contained 2 practice items to familiarize students with the question and response formats, and 10 scored problems. All students answered both multiple-choice and constructed-response problems. The students were randomly assigned to one of two groups in which the question context was altered. Version 1 contained typical word problems, while Version 2 contained problems arranged in an abbreviated list structure.

Results

Scores from the abbreviated list group ($M = 55.45$, $SD = 23.82$, $n = 11$) slightly exceeded the word problem group ($M = 55.00$, $SD = 20.29$, $n = 14$). Response times for the abbreviated list group ($M = 24.66$, $SD = 6.22$, $n = 11$) also exceeded the word problem group ($M = 23.71$, $SD = 6.22$, $n = 14$).

With regard to response format, multiple-choice scores exceeded constructed-response scores, as shown in Table C1. Individual question means for the abbreviated list group exceeded those of the word problem group, with the exception of three questions. (See Table C2.) No discernible pattern between contexts was noted with respect to response time, with the exception of question 1. (See Table C3.)

Table C1

Response Format Average Scores

	Group	M	SD	n
Multiple-Choice	Word	37.14	10.69	14
	List	32.73	16.79	11
	Total	35.20	13.58	25
Constructed-Response	Word	17.86	10.51	14
	List	22.73	11.04	11
	Total	20.00	10.80	25

Table C2

Question Average Scores

Question	Context	M	SD	n
1	Word	.55	.52	14
	List	.09	.30	11
2	Word	.36	.49	14
	List	.45	.52	11
3	Word	.36	.49	14
	List	.27	.47	11
4	Word	.88	.33	14
	List	.91	.30	11
5	Word	.68	.48	14
	List	.82	.40	11
6	Word	.56	.51	14
	List	.55	.52	11
7	Word	.44	.51	14
	List	.45	.52	11
8	Word	.48	.51	14
	List	.54	.52	11
9	Word	.52	.51	14
	List	.73	.47	11
10	Word	.72	.46	14
	List	.73	.47	11

Table C3

Question Average Response Times

Question	Context	M	SD	n
1	Word	29.71	23.35	14
	List	87.91	61.33	11
2	Word	141.71	136.04	14
	List	159.18	116.08	11
3	Word	73.14	29.35	14
	List	74.55	28.62	11
4	Word	129.79	61.25	14
	List	143.55	59.21	11
5	Word	121.21	46.75	14
	List	128.09	44.07	11
6	Word	140.71	74.61	14
	List	117.00	54.29	11
7	Word	99.79	51.72	14
	List	92.91	70.45	11
8	Word	305.93	218.52	14
	List	289.64	134.99	11
9	Word	227.79	73.60	14
	List	227.82	66.61	11
10	Word	152.93	100.17	14
	List	159.00	69.03	11

Discussion. Unlike the adult students from the developmental tryout, the high school students were strongly encouraged by their instructors to "show up well." All students appeared well prepared for the quiz. It is likely that the high school instructors were informed by the adult class instructor concerning the nature of questions on the quiz.

As a result of a program flaw in the computer scoring feature, it was necessary to scrutinize test scores and make appropriate adjustments for rounding errors. The program was subsequently modified to recognize acceptable tolerances before additional data collection.

Student notes were examined closely for patterns in problem solution. Common problem solving strategies included constructing numerical tables, equations, and drawing figures. Students were more likely to represent weight and balance problems in a table structure similar to the tables used in the Aviation Weight and Balance Handbook. The mnemonic device "W A M," symbolizing "Weight X Arm = Movement," occurred frequently on student-constructed tables. Errors that occurred in these problems were examined closely for possible flaws in problem representation (incorrect schema).

Many students constructed line diagrams that visually represented the aircraft with the various weights distributed along a continuum. Several problems were solved by the use of algebraic equations. No obvious differences in how students represented problems were noted between question context and response format.

The students were less likely to represent the first three problems on the test. These were general physics and math problems. These questions were more difficult for the students. The students were more likely to represent correctly the solution to the more complex weight and balance problems.

Question 1 on the abbreviated list version was considerably more difficult than the corresponding word problem. While the problem presented in the two versions is similar, the abbreviated list contains an additional proposition. The word problem may have been easily remembered from the published test bank; however, the underlying principle did not transfer when the same problem was presented in the abbreviated list format. Another explanation is that without contextual information the students did not recognize it as a weight and balance problem. Rearranging questions on the quiz should remove any possible order effect.

Only one student from each group used notepaper to solve this problem. Both students represented the problem correctly; however, there was an arithmetic error in the abbreviated list problem solution. It is unclear why other students failed to solve the problem. Perhaps the abbreviated list problem appeared to be too easy, and students did not bother to check their work for accuracy.

Table C4

Problem Solving Strategies

Id	Ver	P1	P2	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	1			4			3	3	2	2		3	3
2	2							3				3	3
3	1					4	3	3	1,2	1,2	1,3	3	3
4	2						1,2	3	1,2	1,2	3	3	3
5	1						3	3	1,2	1,2			
6	2						3	3	5	1	3		
7	1						3	3				3	3
11	1						3	3		3	3		
12	2		2			5	3	3			3	3	3
13	1									1,2		1,3	
14	2						3	3	1,2		3	3	
15	1	1			1		3	3		5		3	
16	2	1					3	3	1	1	3	3	3
17	1						3	3	1	1	3	3	3
22	2						3	3	2	2	3	3	
23	1	1	1,2				3	3	1,2	1,2			
30	2							3	2,3	2	3	3	3
31	1							3	1		3	3	
32	2					5	3		5		3	3	3
33	1	1	1				3	3	2	1,2	2,3	1,2	3
34	2			1			3	3	1,2		3	3	3
35	1	1	1				3	3	3	5	3	3	3
37	1	1	1						2	2	3	3	3
42	2										3	3	3
45	1	1		1			3	3	2	2	3	3	3

Note. Key: 1 = Figure; 2 = Equation; 3 = Table; 4 = Omit; 5 = Other

Appendix D

AVQUIZ

Table D1

Instructions for Using AVQUIZ

1. You may use calculators during the quiz.
2. You will be provided blank scratch paper during the quiz. Please return to test proctor when finished with the quiz.
3. Click on the AVQUIZ icon on the desktop.
4. Enter the following password: ____________
5. Enter the following Student User ID: ____________
6. Please check the appropriate blocks:
   - Age
   - Gender
   - Education
   - Prior Math/Science Courses
   - Experience
   - English Proficiency
7. Estimate your effectiveness in answering quiz items.
8. Answer the 2 Sample Test Items
   - Multiple-choice format – use mouse to select correct or best answer.
   - Completion – use numeric keypad to enter answer (round to 2 decimal places).
   - Click next to go to next test item.
9. Begin the Quiz.

   (12 questions on aviation weight and balance subject area)
   - You will see the question only once.
   - You will receive no credit for questions you do not answer.

Table D2

Preliminary Information

Version

- (1)Word Problem (MC/CR)
- (2) Abbreviated List Problem (MC/CR)
- (3) Word Problem (CR/MC)
- (4) Abbreviated List Problem (CR/MC)
-

Student ID

Student generated

Age

- 15-17
- 18-21
- 22-25
- 26-29
- 30+

Gender

- Male
- Female

Education

- High school student
- High school graduate
- GED equivalent
- 1-2 yrs college
- College graduate

Prior Math/Science (check all that apply)

- Algebra
- Geometry
- Trigonometry
- Calculus
- Physics

Table D2 (continued)

<u>Preliminary Information</u>

Prior Aviation Experience

- None
- Civilian
- Military

Other ________________

Is English your native language?

- Yes
- No

Please rate your English proficiency (ability to read, write, and speak), using the following scale:

- Very poor
- Fair
- Good
- Very Good
- Excellent

Table D3

Self-Efficacy Questionnaire (Word Problems)

People differ in how confident they are that they can do things. This questionnaire asks you to indicate how confident you are that you could solve aviation weight and balance problems. Before rating the problems, look at an example first.

Example Item

Bob is an excellent basketball player. He was a starter on his high school team. If we gave Bob a questionnaire that asked him how confident he was that he could make a lay-up shot, Bob would indicate he was very confident. On the scale below that corresponds to the number 5. If we asked Bob how confident he was that he could fly an airplane, which he has never done, he would indicate he had little confidence. On the scale below that corresponds to number 1.

Very little ←--→Very high

CONFIDENCE

1 2 3 4 5

Aviation Weight and Balance Problems: For each item below indicate how confident you are that you could solve a problem like this. Use the five-point scale above where 1 = very little confidence and 5 = very high confidence.

Problems	Confidence
A rectangular shaped fuel tank measures 36 inches in depth, 12 inches in width, and 6 ½ inches in depth. How many cubic inches are within the tank?	1 2 3 4 5
An aircraft as loaded weighs 3,375 pounds at a CG of +30.5 inches. The CG range is +32.0 inches to +42.1 inches. Find the minimum weight of the ballast necessary to bring the CG within the CG range. The ballast arm is +180 inches.	1 2 3 4 5
As weighed, the total empty weight of an aircraft is 5,995 pounds with a moment of 885,997 pound-inches. However, when the aircraft was weighed, 24 pounds of potable water were on board at 86 inches, and 18 pounds of hydraulic fluid were in a tank located at 125 inches. What is the empty weight center of gravity of this aircraft?	1 2 3 4 5
An aircraft with an empty weight of 1,500 pounds and an empty weight CG of +28.4 was altered as follows: two 12-pound seats located at +77 were removed; structural modifications weighing +28 pounds were made at +73; a seat and safety belt weighing 30 pounds were installed at +71; and radio equipment weighing 25 pounds was installed at +95. What is the new empty weight CG?	1 2 3 4 5

Table D4

Self-Efficacy Questionnaire (Abbreviated List)

People differ in how confident they are that they can do things. This questionnaire asks you to indicate how confident you are that you could solve aviation weight and balance problems. Before rating the problems, look at an example first.

Example Item

Bob is an excellent basketball player. He was a starter on his high school team. If we gave Bob a questionnaire that asked him how confident he was that he could make a lay-up shot, Bob would indicate he was very confident. On the scale below that corresponds to the number 5. If we asked Bob how confident he was that he could fly an airplane, which he has never done, he would indicate he had little confidence. On the scale below that corresponds to number 1.

Very little ←--→Very high

CONFIDENCE

1	2	3	4	5

Aviation Weight and Balance Problems: For each item below indicate how confident you are that you could solve a problem like this. Use the five-point scale above where 1 = very little confidence and 5 = very high confidence.

Problems	Confidence
Given rectangular fuel tank: Length = 36 in Width = 12 in Depth = 6 1/2 in Volume = ?	1 2 3 4 5
Given: Aircraft EW = 3,375 lbs CG = +30.5 in CG range is +32.0 to +42.1 in Ballast @ +180 in Ballast wt = ?	1 2 3 4 5
Given: Aircraft EW = 5,995 lbs Moment = 885,997 Potable water = 24 lbs @ +86 Hydraulic fluid = 18 lbs @ +125 EWCG = ?	1 2 3 4 5

Table D4 continued

Self-Efficacy Questionnaire (Abbreviated List)

Problems	Confidence
Given: Aircraft EW = 1,500 lbs @ +28.4 Remove: Two seats = 12 lbs @ +77 Install: Structural mods = 28 lbs @ +73 Seat/safety belt = 30 lbs @ +71 Radio equip = 25 lbs @ +95 New EWCG = ?	1 2 3 4 5

Table D5.

<u>Version 1 – Word Problems</u>[1]

<u>Practice Questions</u>

1. In the rectangle ABCD, determine the area of the triangle formed by points A, B, and C.

   AC = 7.5 inches
   CD = 16.8 inches

   - 42 square inches
   - 63 square inches
   - 126 square inches

2. Compute the area of a trapezoid whose altitude is two feet and the length of each side is four feet and six feet respectively.

[1] Order of multiple-choice and constructed-response items is reversed in Version 3.

AVQUIZ

<u>Directions:</u>

Multiple-choice questions: Circle the letter of the correct answer.

Constructed-response questions: Round your answer to 2 decimal places.

1. How many gallons of fuel will be contained in a rectangular-shaped tank which measures two feet in width, three feet in length, and one foot eight inches in depth? (Note: Seven and one-half gallons = one cubic foot)

   - 66.6 gallons
   - 75 gallons
   - 45 gallons

2. Two boxes that weigh 10 pounds and 15 pounds are placed in an airplane so that their distance aft from the center of gravity are 7 feet and 2 feet respectively. A 5-pound box is placed forward of the center of gravity 3 feet. How far forward of the center of gravity should a fourth box, weighing 20 pounds, be placed so that the center of gravity will not be changed?

   - 3.5 feet
   - 4.25 feet
   - 4.5 feet

3. As weighed, the total empty weight of an aircraft is 5,862 pounds with a moment of 885,957 pound-inches. However, when the aircraft was weighed, 20 pounds of potable water were on board at 84 inches, and 23 pounds of hydraulic fluid were in a tank located at 101 inches. What is the empty weight center of gravity of the aircraft?

   - 150.700 inches
   - 151.700 inches
   - 151.365 inches

4. An aircraft as loaded weighs 2,500 pounds at a center of gravity of 38.4 inches. The center of gravity range of the aircraft is 40.9 to 46.0 inches. Find the minimum weight of the ballast necessary to bring the center of gravity within the given range. The ballast arm is located at 120 inches.

   - 63.21 pounds
   - 69.53 pounds
   - 79.01 pounds

5. An aircraft had an empty weight of 2,886 pounds with a moment of 101,673.78 before several alterations were made. The alterations included: removing two passenger seats weighing 15 pounds each at 71 inches; installing a cabinet weighing 97 pounds at 71 inches; installing a seat and safety belt weighing 20 pounds at 71 inches, and installing radio equipment weighing 30 pounds at 94 inches. What is the new empty weight center of gravity?

   - 33.20
   - 36.85
   - 37.26

6. Given an aircraft where the datum is forward of the main gear center point at 30.15 inches, and the actual distance between tail gear and main gear center points is 360.65 inches. Also the net weight at the right main gear is 9,750 pounds, while the net weight at the left main gear is 9,960 pounds, and the net weight at the tail gear is 1,940 pounds. These items were in the aircraft when weighed: full lavatory water tank weighing 32 pounds at 325 inches, hydraulic fluid weighing 22 pounds at minus eight inches, and removable ballast weighing 150 pounds at 380 inches. What is the empty weight center of gravity of the aircraft?

   - 59.86 inches
   - 60.31 inches
   - 62.44 inches

7. A rectangular-shaped fuel tank measures 37 and one-half inches in length, 14 inches in width, and eight and one-fourth inches in depth. How many cubic inches are within the tank?

8. The total empty weight of an aircraft is 1,800 pounds with a moment of 56,700 pound-inches. When the aircraft was weighed, 18 pounds of potable water were on board at 21.5 inches, and 12 pounds of hydraulic fluid were in a tank located at 24 inches. What is the empty weight center of gravity of the aircraft?

9. An aircraft as loaded weighs 4,954 pounds at a center of gravity of 30.5 inches. The center of gravity range of the aircraft is 32.0 inches to 42.1 inches. Find the minimum weight of the ballast necessary to bring the center of gravity within the given range. The ballast arm is located at 162 inches.

10. Two boxes which weigh 10 pounds and five pounds are placed in an airplane so that their distance from the center of gravity are 4 feet and 2 feet respectively. How far forward of the center of gravity should the third box, weighing 20 pounds, be placed so that the center of gravity will not be changed?

11. Given an aircraft where the datum is forward of the main gear center point at 30.24 inches, and the actual distance between tail gear and main gear center points is 360.26 inches. Also the net weight at the right main gear is 9,980 pounds, while the net weight at the left main gear is 9,770 pounds, and the net weight at the tail gear is 1,970 pounds. These items were in the aircraft when weighed: full lavatory water tank weighing 34 pounds at 352 inches, hydraulic fluid weighing 22 pounds at minus eight inches, and removable ballast weighing 146 pounds at 380 inches. What is the empty weight center of gravity of the aircraft?

12. An aircraft with an empty weight of 2,100 pounds and an empty weight center of gravity of 32.5 inches was altered as follows: two 18-pound passenger seats located at 73 inches were removed; structural modifications were made at 77 inches increasing weight by 17 pounds; a seat and safety belt weighing 25 pounds were installed at 74.5 inches; and radio equipment weighing 35 pounds was installed at 95 inches. What is the new empty weight center of gravity?

Table D6.

Version 2 – Abbreviated List Problems[2]

Practice Questions

1. Given right triangle:
   Altitude = 7.5 in
   Base = 16.8 in
   Area = ?

- ○ 42 in^2
- ○ 63 in^2
- ○ 126 in^2

2. Given trapezoid:
   Altitude = 2 ft
   Base 1 = 4 ft
   Base 2 = 6 ft
   Area = ?

[2] Order of multiple-choice and constructed-response items is reversed in Version 4.

AVQUIZ

Directions:

Multiple-choice questions: Circle the letter of the correct answer.

Constructed-response questions: Round your answer to 2 decimal places.

1. Given rectangular fuel tank:
   Length = 3 ft
   Width = 2 ft
   Depth = 1 ft 8 in
   Volume = ? (Note: 7 ½ gal = 1 ft^3.)

   - 66.6 gal
   - 75 gal
   - 45 gal

2. Given aircraft:
   10 lb box @+7
   15 lb box @ +2
   5 lb box @ -3
   20 lb box @ ?
   (Assume CG does not change.)

   - -3.5
   - -4.25
   - -4.5

3. Given:
   Aircraft EW = 5,862 lbs
   Moment = 885,957
   Potable water = 20 lbs@+84
   Hydraulic fluid = 23lbs@+101
   EWCG = ?

   - +150.700
   - +151.700
   - +151.365

4. Given:
   Aircraft EW = 2,500 lbs @ +38.4
   CG range = +40.9 to +46.0
   Ballast @ +120 in
   Ballast wt = ?

   - 63.21 lbs
   - 69.53 lbs
   - 79.01 lbs

5. Given:
   Aircraft EW = 2,886 lbs
   Moment = 101,673.78
   Remove: Two seats = 15 lbs each @ +71.
   Install: Cabinet = 97 lbs @ +71;
   Seat/safety belt = 20 lbs @ +71;
   Radio equip.= 30 lbs @ +94.
   New EWCG = ?

   - +33.20
   - +36.85
   - +37.26

6. Given:
   Configuration as weighed
   Right gear = 9,750 lbs @ +30.15
   Left gear = 9,960 lbs @ +30.15
   Tail gear = 1,940 lbs @ +390.5
   Full lavatory tank = 32 lbs @ +325
   Hydraulic fluid = 22 lbs @ -8
   Temporary ballast = 150 lbs @ + 380
   EWCG = ?

   - +59.86 in
   - +60.31 in
   - +62.44 in

7. Given rectangular fuel tank:
   Length = 37 ½ in
   Width = 14 in
   Depth = 8 ¼ in
   Volume = ?

8. Given:
   Aircraft EW = 1,800 lbs
   Moment = 56,700
   Potable water = 18 lbs @ +21.5
   Hydraulic fluid = 12 lbs @ +24
   EWCG = ?

9. Given:
   Aircraft EW = 4,954 lbs @ +30.5
   CG range = +32 to +42.1
   Add: Ballast @ +162
   Ballast wt = ?

10. Given aircraft:
    10 lb box @ +4
    5 lb box @ +2
    20 lb box @?
    (Assume CG does not change)

11. Given:
    <u>Configuration as weighed</u>
    Right gear = 9,980 lbs @ +30.24
    Left gear = 9,770 lbs @ +30.24
    Tail gear = 1,970 lbs @ +390.5
    Full lavatory tank = 34 lbs @ +352
    Hydraulic fluid = 22 lbs @ -8
    Temporary ballast = 146 lbs @ + 380
    EWCG = ?

12. Given:
    Aircraft EW = 2,100 lbs @ +32.5
    Remove: Two seats = 18 lbs each @ +73.
    Install: Struct. Mods.= 17 lbs @ +77;
    Seat/safety belt = 25 lbs @ +74.5;
    Radio equip. = 35 lbs @ +95.
    EWCG = ?

Appendix E

Consent Form for Participants in a Research Project
under the auspices of
the University of Oklahoma, Norman Campus

You are being asked to participate in a study to evaluate the difficulty of different types of aviation weight and balance problems. To participate, you must be 16 years of age or older. The study is being conducted by Linda McCoy, doctoral student in the Instructional Psychology and Technology program. The purpose of this study is to determine whether one type of test item is more difficult than another.

Today you will complete a computer-administered achievement test on the topic of weight and balance computations. Your responses will be used to determine which types of test items are most difficult in solving aviation weight and balance problems.

You will also be asked to answer some questions about the types of test items. The whole process should take approximately one hour. You will not be taking any risks or be harmed by this research.

Your participation is voluntary. You can stop at any time and will not be penalized in any way. To make sure responses are confidential, your name will not go on the computer test form you will complete.

If you have questions about your rights as a research participant, you may contact the Office of Research Administration at 405-325-4757 or email: irb @ ou.edu. If you have any questions about this study, you may contact Linda McCoy at 405-954-6401

I agree to participate in this study

I understand all of the above statements.

Name ____________________

Date ____________________

Parental Consent Form for Participants in a Research Project
under the auspices of
the University of Oklahoma, Norman Campus

Your son/daughter is being asked to participate in a study to evaluate the difficulty of different types of aviation weight and balance problems. To participate, they must be 16 years of age or older. The study is being conducted by Linda McCoy, doctoral student in the Instructional Psychology and Technology program. The purpose of this study is to determine whether one type of test item is more difficult than another.

They will complete a computer-administered achievement test on the topic of weight and balance computations. Their responses will be used to determine which types of test items are most difficult in solving aviation weight and balance problems.

They will also be asked to answer some questions about the types of test items. The whole process should take approximately one hour. Your son/daughter will not be taking any risks or be harmed by this research.

Participation is voluntary. They can stop at any time and will not be penalized in any way. To make sure responses are confidential, their name will not go on the computer test form they will complete.

If you have questions about the rights of your son/daughter as a research participant, you may contact the Office of Research Administration at 405-325-4757 or email: irb @ ou.edu. If you have any questions about this study, you may contact Linda McCoy at 405-954-6401.

I agree to allow my son/daughter to participate in this study.

I understand all of the above statements.

____

Name

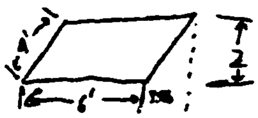
____

Date

Appendix F

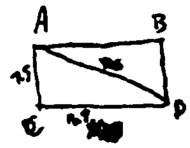
Student Notes

The following pages contain examples of student notes collected during the administration of AVQUIZ at the four schools. Examples range from the minimum score attained on the quiz (8 percent) to the maximum score (100 percent).



Area = $\left(\frac{A+B}{2}\right)H$ = $\frac{(8+6)}{2}(2)$ = 29 ft²

round 2 places



$A = \frac{1}{2}(11.4)(2.5) = 63.2$

① $V = (37.5)(14)(8.25) = 4331.25$ in²

② $0 = -10(4) - 5(2) + 20(x)$ ∴ $x = 2.5$

③ EW = 1800 lb − water = 1800 − 18 = 1782 lb   CG = 31.60

Moment = 56,700 lb·in − (18)(21.5) = 56313

④ Weight = 4954 ∴ x = 57.16 lb

CG = 30.5"

Range = 32 − 42.1 in

Ballast arm = 162"

$$32 = \frac{(4954)(30.5) + (x)(162)}{4954 + x}$$

⑤ EW = 2100 lb   Moment = 68250

EWCG = 32.5

New weight = 2100 − 2(18) + 17 + 25 + 35 = 2141

New Moment = 68250 − 2(18)(73) + 17(77) + 25(79.5) + 35(95) = 72118.5

New EWCG = 33.68

Remove	Added
(2) 18 lb @ 73"	17 lb @ 77"
	25 lb @ 79.5"
	35 lb @ 95"

⑥ Rt. gear − 9980 @ 30.24"   Remove

Lt. gear − 9770 @ 30.24"   34 lb @ 352"

tail − 1970 @ 390.5"   96 lb @ 380"

total wt = 21590

Moment = (9980)(30.24) + (9770)(30.24) + (1970)(390.5) − (34)(352) − (96)(380) = 1.2991 × 10⁶ lb·in

CG = 60.31

⑦ $N = (2)(3)(1.167) = (\ ft^3)(7.48)$ =

⑧ EW = 5862   remove − 20 lb @ 84" $CG = \frac{885957 - 20(84)}{5862 - 20} = 151.71$

Moment = 885,957 lb·in

Figure E1. Student Exhibit 1.

9) W = 2500 lb
CG = 38.7 in
Range = 40.9 - 46.0

$$40.9 = \frac{(2500)(38.7) + x(120)}{2500 + x} \quad \therefore x = 79.01$$

10) $0 = -10(7) - 15(3) + 5(3) + 20(x) \quad \therefore x = 4.25$

11) Rt → 9750 lb @ 30.15
Lt → 9960 lb @ 30.15
tail → 1970 lb @ 392.8

Remove
32 @ 325
150 @ 380

Wt = 21 468
CG = 59.86
Moment = 1.285×10^6

12) EW = 2886
Mom. = 101 673.78
CG = 36.85

Remove
2(15) @ 71"

Add
97 lb @ 71"
20 lb @ 71"
30 lb @ 97"

Figure E1. Student Exhibit 1 (continued).

$$\frac{M + \Delta M}{W + \Delta W} = CG$$

$3 \; \boxed{1.84} \; 10^{13}$

2

1) $10 \times 7 = 70$
$15 \times 2 = 30$
$5 \times -3 = -15$

30 $\quad$ 85

$\cancel{(10\times4)+(5\times2)}$

$$\frac{40 + 10 + (20 \times X)}{\cancel{70}} = 0$$

$10 + 5 + 20$

$$\frac{885{,}957 - (20 \times 84) - (23 \times 101)}{5{,}062 - 20 - 23}$$

$$\frac{85 + 20x}{50} = 0$$

$\cancel{2500}$

$\cancel{\frac{(2500 + X) \cdot 40.9}{2500 + X}} = 40.9$

$\cancel{101{,}673.78 \times (19 \times 71)}$

$$\frac{101{,}673.78 - (30 \times 71) + (97 \times 71) + (20 \times 71) + (30 \times 94)}{2986 - 30 + 97 + 20 + 30}$$

$$\frac{(2400 \times 38.4) + (x \times 120)}{2500 + x} = 40.9$$

tail arm = 390.8"
main arm = 30.15"
mains = 19710
tail = 1940

$$\frac{(19710 \times 30.15) + (1940 \times 390.8)}{19710 + 1940} = 62.467$$

$$\frac{56{,}700 - (18 \times 215)}{1800 - \cancel{215}\ 18}$$

$$\frac{1{,}352{,}408.5 - (32 \times 325) - (150 \times 380)}{21650 - 32 - 150} = 59.86$$

$$\frac{151097 + (x \cdot 162)}{4954 + x} = 32.0$$

$14 \; \boxed{8.25A}$

37.5

Figure E2. Student Exhibit 2.

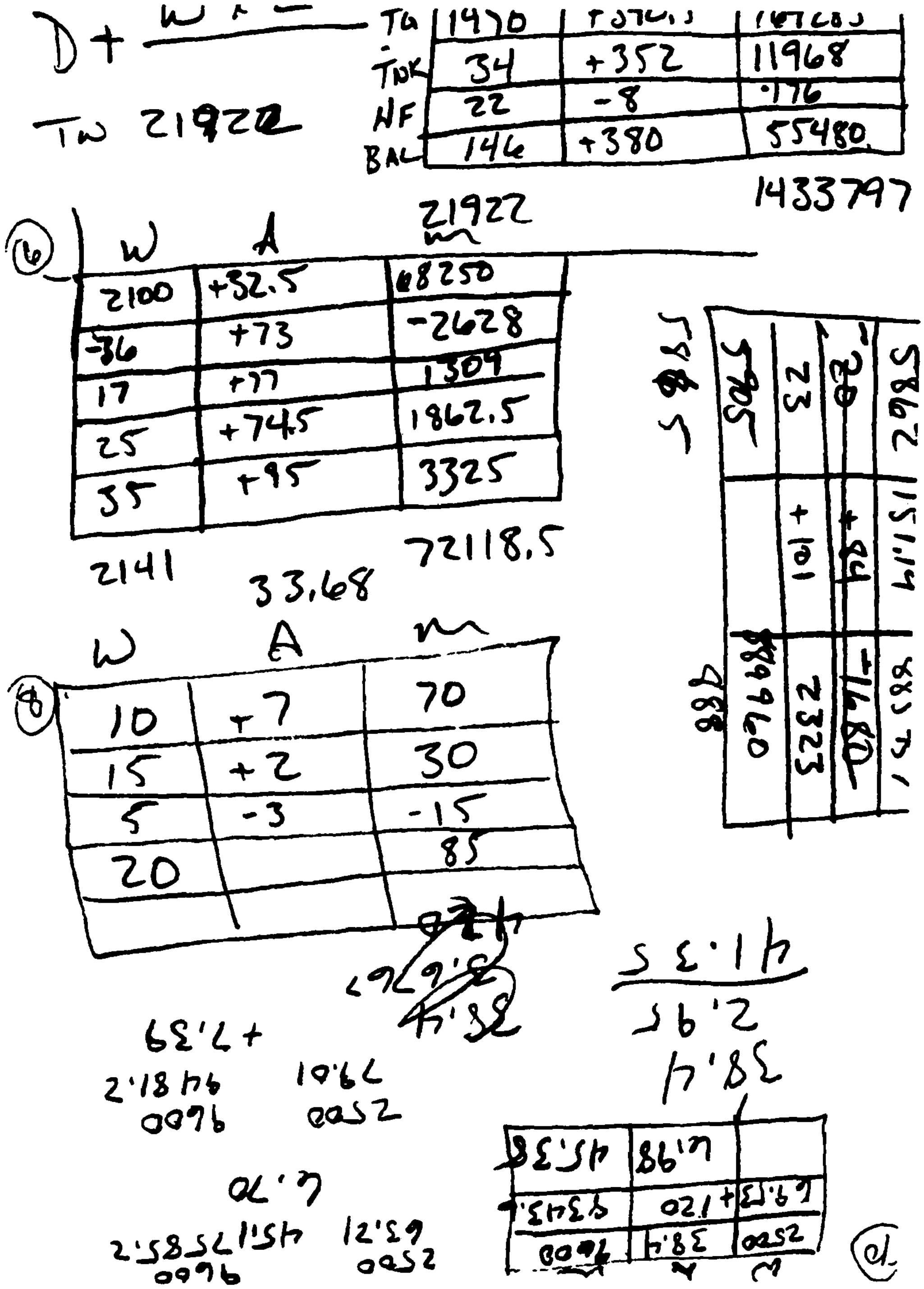


Figure E3. Student Exhibit 3.

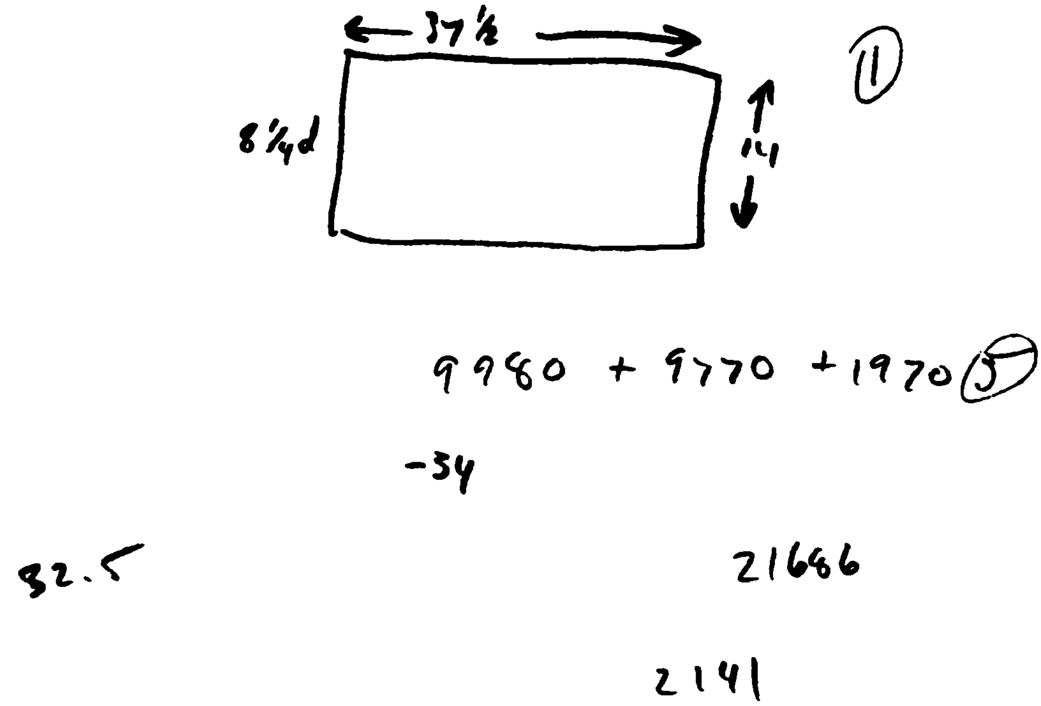


Figure E4. Student Exhibit 4.

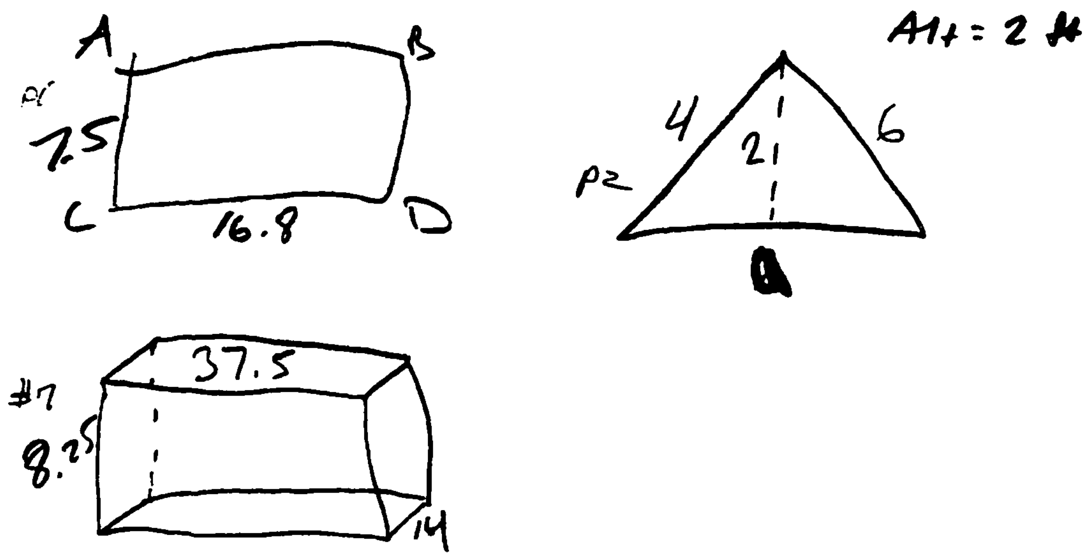


Figure E5. Student Exhibit 5.

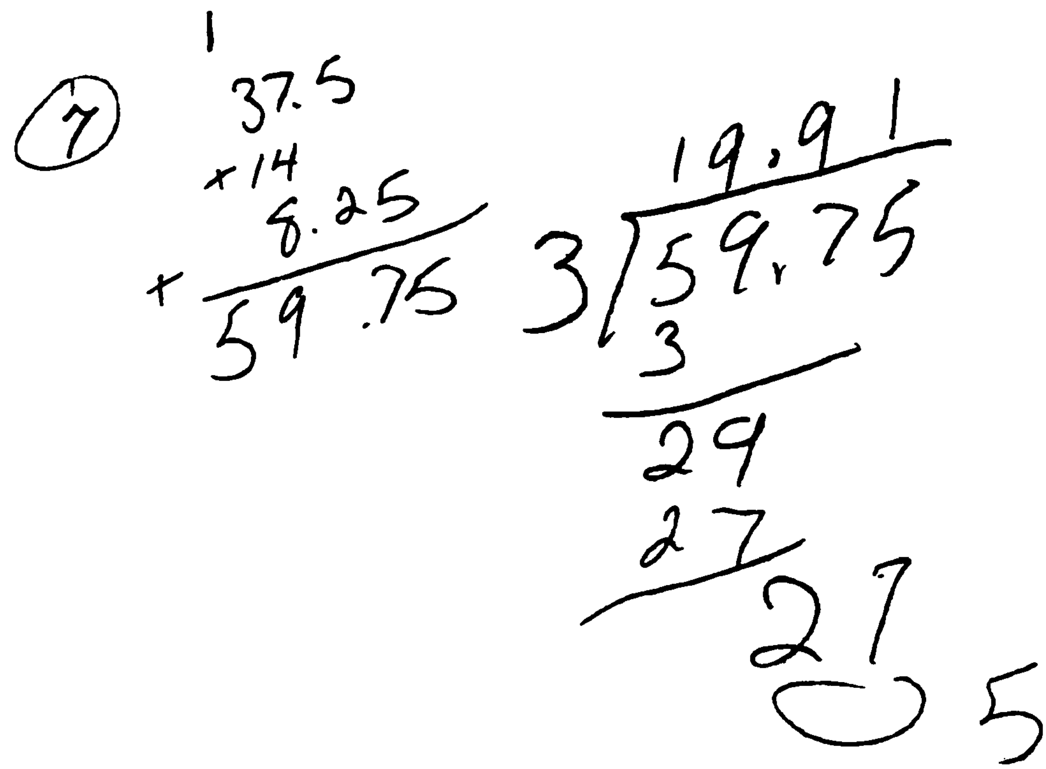


Figure E6. Student Exhibit 6

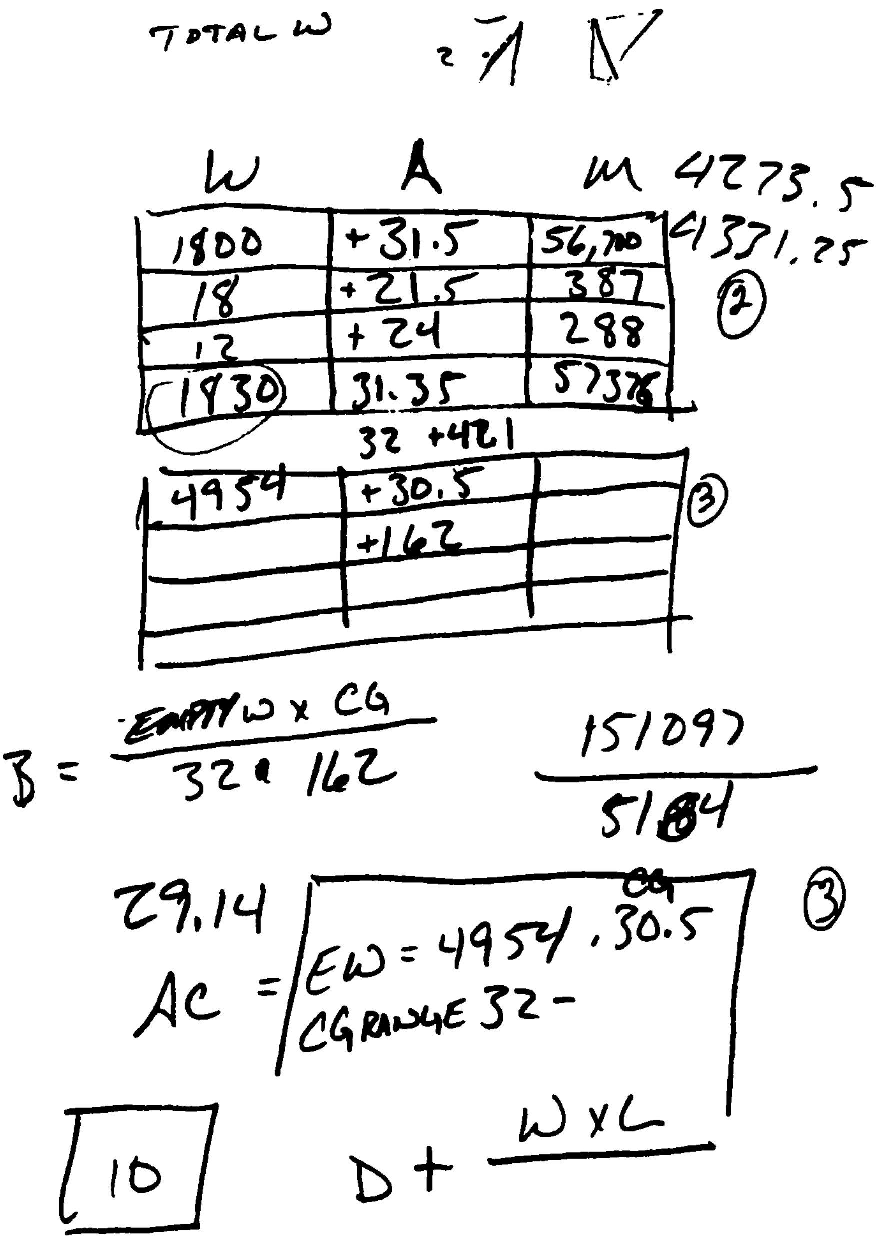


Figure E7. Student Exhibit 7.

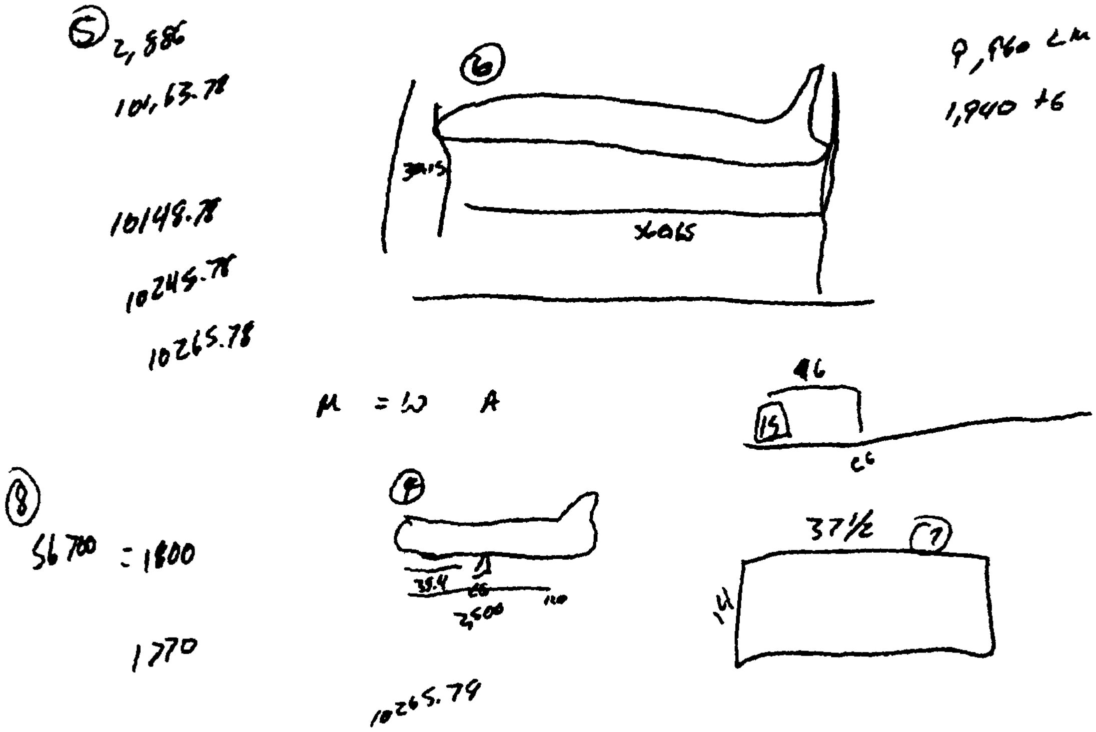


Figure E8. Student Exhibit 8.

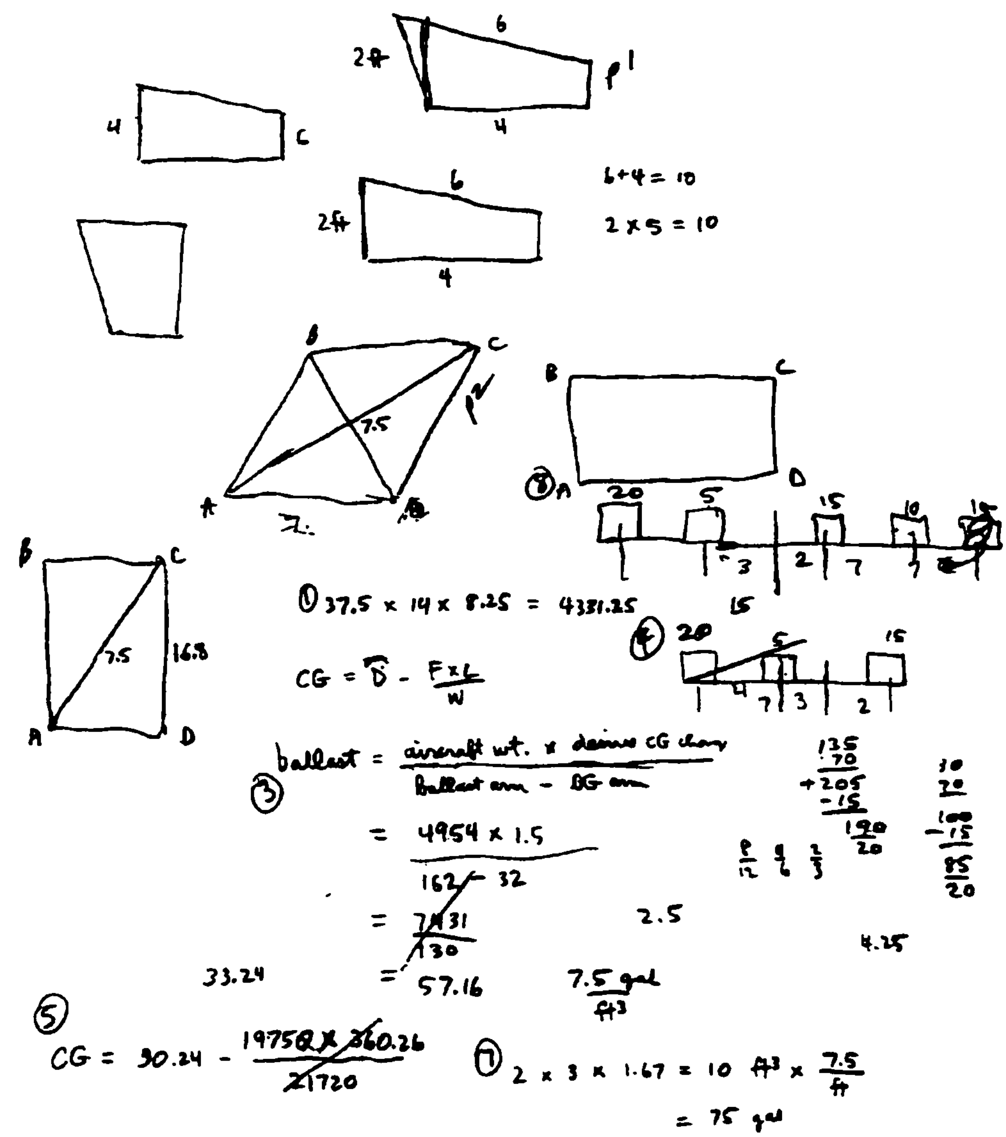


Figure E9. Student Exhibit 9.

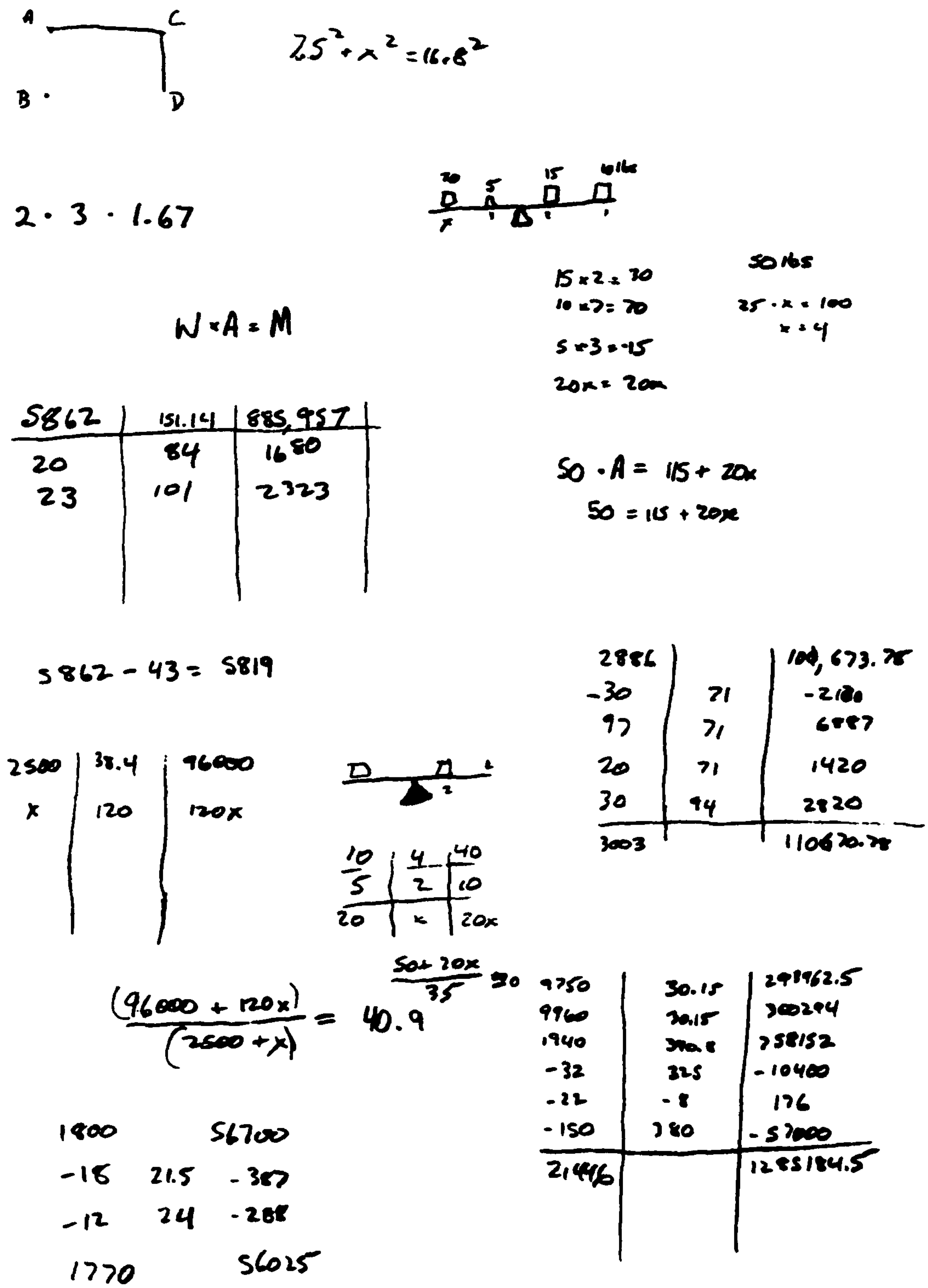


Figure E10. Student Exhibit 10.

9980	3024	301795.2
9770	30.24	295444.8
1970	390.5	769285
-34	352	-13277
-22	-8	176
-146	300	-55480
		1297944

2100	32.5	68250
-36	73	-2628
17	77	1309
25	74.5	1862.5
35	95	3325
2141		72118.5

Figure E11. Student Exhibit 11.

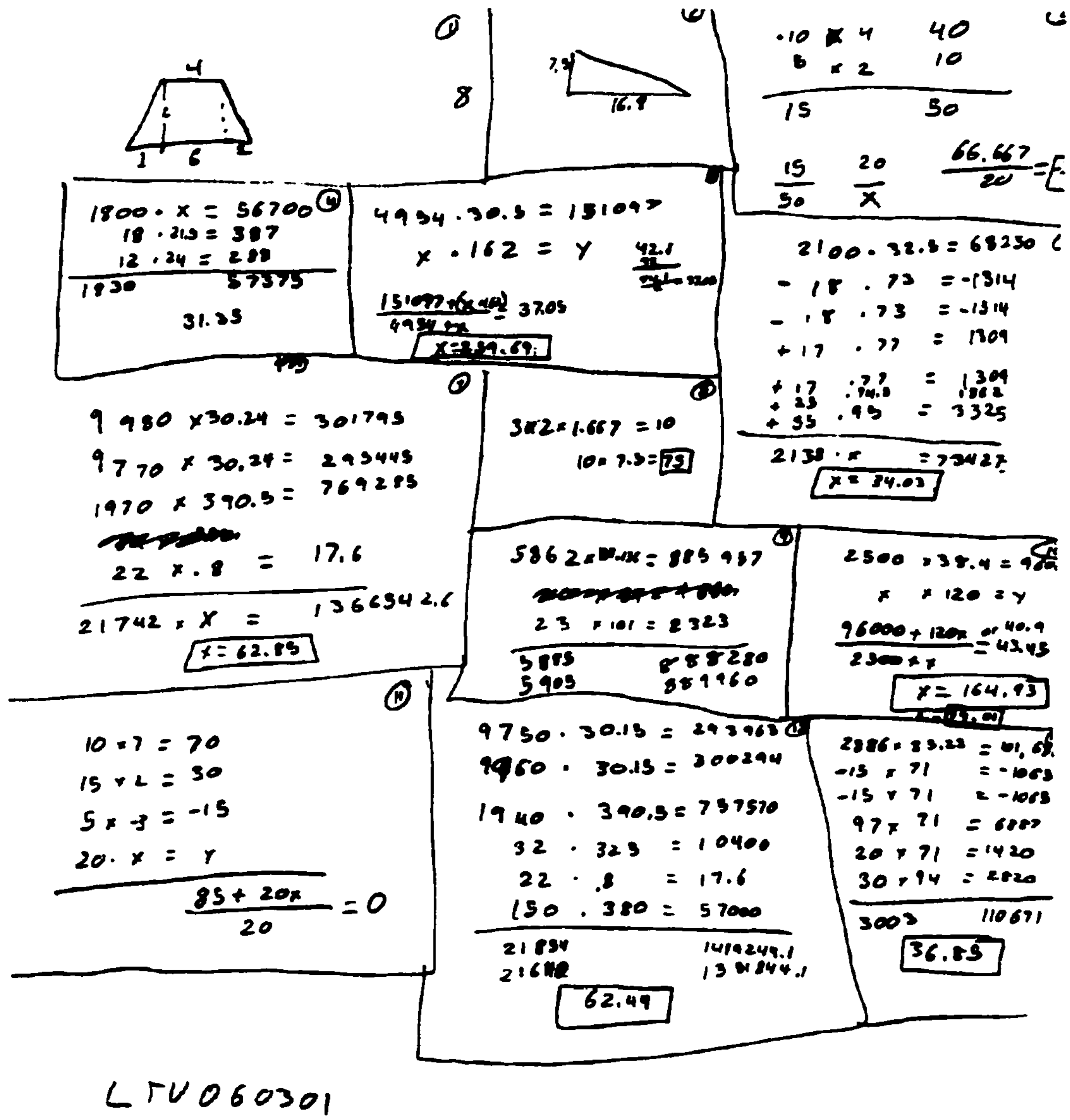


Figure E12. Student Exhibit 12.