Assessment of the Long-Term Effects of Technology Use in the Engineering Classroom on Learning and Knowledge Retention

A Thesis
Presented to
The Academic Faculty

By

Sean St.Clair

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
April 2004

Assessment of the Long-Term Effects of Technology Use in the Engineering Classroom on Learning and Knowledge Retention

Approved by:

Dr. Nelson Baker, Advisor	Dr. Lawrence Jacobs
Dr. Reginald DesRoches	Dr. Joseph Hoey
Dr. David Elrod	Dr. M. Jackson Marr

March 24, 2004

This work, of course, is dedicated to my beautiful wife and children. Without them, all my efforts would be meaningless.

ACKNOWLEDGEMENTS

There are a number of people and organizations without whom this work would have been impossible. I would like to acknowledge SUCCEED and NSF for the financial support that has been provided over the years; I would like to think that the money all went to a good cause.

I would like to thank my advisor, Dr. Nelson Baker, for his incessant support and wisdom. He always encouraged me to be better than I was. I am also grateful that he was never apologetic about the type of research that we do, despite the opinions and concerns of others. I am proud to have worked with him and look forward to many future collaborations.

I would also like to thank the members of my committee; Drs. Jacobs and DesRoches for revealing to me the intricacies and limitations of conducting this work in a statics class, Dr. Elrod for his unique perspective on educational research, Dr. Hoey for constantly reminding me of the importance of this study, and Dr. Marr, whose mere presence on my committee intimidated me sufficiently to make sure that I did it right the first time.

My gratitude also goes out to the instructors and students who participated in the experiments, I could not have done this without their wholehearted participation and support.

Many thanks are also due to all my EPITOME colleagues who were always willing to give their honest opinions. I would like to especially thank Tulio Sulbaran and

Ahmet Citipitioglu for their friendship and support. Graduate school was a whole lot more fun with the two of them around.

I would like to thank Ralph Kern and Ralph Francis. They unknowingly and unintentionally taught me a very valuable lesson seventeen years ago. They taught me to never give up, and that I can do anything I put my mind to. I draw strength from that experience to this day and will continue to for the rest of my life.

Appreciation also goes out to my family. I am who I am because of the love, support, and encouragement of my parents. They are amazing people and I can only hope to be the kind of parent that they both are. I am also grateful to my five siblings, all of whom have grown up to become amazing people and good friends.

I would, of course, like to thank my Father in Heaven. He has blessed me, and continues to bless me, with so much. I often feel unworthy and undeserving, but He continues to give me all that I need and more.

Lastly, and most importantly, I would like to thank my amazing wife and beautiful children. Zauhna's support has never wavered. She is a wonderful wife and mother and I look forward to spending eternity with her. She is the reason for everything that I do and the answer to all my prayers. I am nothing without her.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

A longitudinal study of the effects of instructional technology on learning and knowledge retention was conducted in the School of Civil and Environmental Engineering at Georgia Tech. Instructional technology has been promoted as a means of improving knowledge retention among engineering students. The practical, long-term effects of such technology use were assessed at numerous times over a period of twenty-five weeks. Students in various sections of an undergraduate mechanics course used two different software titles, a structural analysis tool and an electronic textbook, in their studies of trusses and truss analysis. Two other sections of the same course used no software in their classes but spent class time solving problems by hand in teams. All sections were taught truss analysis by the same guest lecturer who also facilitated in the intervention. Demographic data, including gender, ethnicity, grade point average, and course load, were gathered from each of the sections and compared to assure group equality. Pretests were completed by students in each of the sections and also compared among treatment groups to assure that all sections had equivalent levels of prior knowledge. All students were tested immediately after the intervention to assess their learning of the material. Students were again tested ten and twenty-five weeks after the intervention to assess their long-term retention of the material. Results indicated that technology use increased students' problem solving efficiency. The results of the assessments further indicated that all students had high levels of knowledge retention, but that there were no differential levels of learning or retention among the different groups.

It was thus concluded that instructional technology can make the educational process more efficient without hindering long-term knowledge retention. It was further concluded that solving problems by hand in teams was just as effective at leading to high levels of performance over time as using instructional technology.

CHAPTER 1

INTRODUCTION

1.1 Research Overview

The purpose of this study was to investigate a possible link between instructional technology and long-term retention of engineering knowledge. As described in the following chapter, research has shown that knowledge retention is poor in higher education, particularly in the field of engineering, and that alternative instruction strategies, such as technology implementation, could potentially improve knowledge retention. There are, however, few longitudinal studies conducted in the classroom to support the theorized relationship between instructional technology and knowledge retention. As such, this study was designed to further define that relationship. There is also very little literature on the effects resulting from the application of different types of instructional technology and so this research was designed to add to that knowledge base as well.

The resulting objectives of this research were as follows:

- To determine the effects of instructional technology on learning, retention, and long-term retention.
- To determine whether the effects of content-type software on learning, retention, and long-term retention are different from the effects of tool-type software on these same outcomes.

To accomplish these objectives, an experiment was designed and implemented to longitudinally assess the effects of these two types of technologies (which are defined in the following chapter). Seminal research in the area of knowledge retention in education recommended that studies be conducted in naturalistic settings (i.e. the classroom). This suggestion was applied in this research and the effects of technology were studied in three statics classes taught at Georgia Tech. Each of the two types of technology were implemented in the truss analysis portion of a section of statics. Another section of statics was included in the study as the control group; this section did not use software but relied upon traditional instructional techniques. Because entire sections of the statics course served as the study groups, the experiment was quasi-experimental and thus subject to a potentially confounding selection bias. To overcome this bias, the experiment was conducted twice, in subsequent semesters.

The sections were assessed at various points in time to determine their relative degrees of knowledge. A pretest was administered prior to the intervention to assess students' prior knowledge. The students were again assessed, via a posttest, after the intervention to measure learning. Ten weeks after the intervention, students again completed the posttest as a measure of retention. The final assessment was conducted at 25 weeks, where a sample of the students once again completed the posttest to measure long-term retention. These research intervals of 10 and 25 weeks were chosen, in keeping with the theme of naturalistic settings, because they represent the amount of time between the intervention and the end of the semester and the length of time between the intervention and a point in time in the middle of the subsequent semester respectively. Instructors assume that students will at least retain information until the end of the semester and, ideally, well into subsequent semesters. As such, these time frames were naturalistic as was the classroom setting in which the experiments were conducted.

The results of these assessments were compared across treatment conditions to determine if computer use had an effect on learning and retention. Comparisons were also made across treatment conditions to determine the effects of different types of technology on performance and retention. These comparisons were made to satisfy the aforementioned research objectives. Additionally, comparisons were made across time to determine the degree to which engineering knowledge was retained.

The results of this research suggest that retention can be improved through instructional technology and other activities. In the final chapters of this document, there are practical suggests and strategies that engineering educators can employ to help improve the retention of engineering knowledge in their students. While engineering educators are the intended audience of this work, instructors in any field will find the results illustrative and practical. This document presents the longitudinal study in its entirety, from the development of the research questions to the suggestions for future work. The organization of this document is described in the following section.

1.2 Organization of Dissertation

This dissertation follows the logical and chronological progression of the study it describes. This section is an outline of the remaining chapters of this document, detailing the entire process of not only the intervention method, results, and findings, but also the review of background literature and software selection process. A brief description of the each of the chapters is given to provide direction to the organization of this document.

Following this introduction is a review of the pertinent background literature. A snapshot of the literature on long-term retention of knowledge is presented. Chapter 2 reveals that long-term retention studies in naturalistic settings are rare but important, and that while retention is poor, there are ways to improve it. Literature presented on instructional technology suggests that technology, when used in the classroom, may increase knowledge retention in engineering education. A definition of instructional technology is followed by examples from engineering education classified according to the two types of technology. The chapter then progresses into the development of the research questions.

The next chapter describes the methodology of the study as it was originally proposed. As the chapter explains, the actual implementation of the intervention changed throughout the course of the study. To illustrate how the nature of the study changed over time, the proposed and the implemented methodologies are both described, but in different chapters, specifically Chapters 3 and 6 respectively. Chapter 3 describes the variables, hypotheses, testing procedures, and intervention plans for the study.

One important part of the method that was not included in Chapter 3 was the software selection. The software selection process was nearly a project in itself and was complicated enough to warrant a separate chapter. As detailed in Chapter 4, the selection process involved the choosing of two software titles that would be used later in the intervention. Educational theory, as it applies to the use of software in the classroom, is briefly reviewed in this chapter. Suggestions from six statics instructors, combined with the suggestions from educational theory, provided the framework used to select the instructional technology titles used in this study.

Chapter 5 describes another project-within-a-project. The formative study presented in Chapter 5 took place in the summer semester prior to the initial implementation of the actual study. The purpose of the formative study was to evaluate the assessment instruments for usability, reliability, and validity. Chapter 5 details the entire process of the formative study and reveals how the results were used to shape and revise the final form of the assessment instruments.

The intervention as it actually took place is described in Chapter 6. This chapter is divided into two parts, the first of which describes what specific changes were made to the proposed methodology as well as why they were made. The second half of the chapter describes in detail what occurred during each of the two implementations of the research process.

The results of each of the assessments are presented in great detail in Chapter 7. This chapter begins with a review of the research hypotheses. A review of the statistical procedures used in the analysis of the data from this project are presented for the benefit of those who are unfamiliar with statistical tests common to behavioral research. Following this review are the results of the data analyses, categorized by assessment instrument and further delineated by the different semesters in which the study was completed. This chapter simply presents the results of the data analyses but does not discuss the meaning of the results.

The findings from the data analyses are discussed in Chapter 8. The hypotheses are once again presented and, in light of the results, are either rejected or retained. Conclusions based on these findings are presented as are practical suggestions for engineering educators. The chapter concludes with areas of suggested future research within the fields of long-term retention and instructional technology.

The final chapter is an extended summary of this document. It briefly describes the entire research process including the results, findings, and conclusions.

In conclusion, this document describes an interdisciplinary project that integrated research from the areas of psychology, cognitive science, instructional technology, education, and structural engineering. This document details every step of the project, including the efforts employed to assure that the experiment met the rigors and standards of each of the contributing fields of study. Furthermore, this document contains a framework that can be followed by future researchers who are interested in conducting longitudinal studies of retention. Practical suggestions for engineering educators who are interested in either implementing technology or increasing their students' retention of knowledge are provided at the end of this dissertation.

CHAPTER 2

BACKGROUND

As stated in the introduction, the objective of this project was to answer the following two research questions:

1. Does the use of instructional technology in the engineering classroom increase long-term knowledge retention?
2. Is there a difference in the long-term effects when using a tool-type software as opposed to a content-type software (these terms are defined below)?

These two questions stemmed from research that has already been conducted in the areas of instructional technology and retention. This literature is presented in this chapter, which is organized according to the following outline. Retention is discussed first including a definition of retention and a snapshot of some of the research that has been completed in this area. Following this is a definition of instructional technology and some typical examples from engineering are presented. The assessment of instructional technology and its impact on retention is presented next with specific emphasis placed on the examples mentioned above. Lastly, the development of the research questions is explained.

Prior to beginning the literature review, however, one point of distinction must be made. Semb and Ellis (1994) have pointed out that there are two general types of retention studies or that most retention studies aim for one of two type of conclusions. The first is functional or practical in nature and tries to determine the effects of specific variables on memory. The second is more theoretical in nature and attempts to describe the internal processes and structure of memory. The first type of research answers the *what* questions (e.g. what is the effect of visualization on retention?). The second type of research answers the *why* or *how* questions (e.g. why do advance organizers effect retention or how do advance organizers relate to specific memory models?).

The work done in this project follows the first school of thought and tried to determine a relationship between IT and long-term retention. This work will not attempt to describe how IT affects mental models or cognitive structures. As such, the focus of this literature review is practical in nature and includes examples of instructional technology, assessment of retention, and descriptions of variables that have been shown to have an effect on retention. Literature in cognitive science and educational theory will be very limited. Some cognitive science research is presented in this chapter as it pertains to practical education and some educational theory is touched upon in Chapter 4 as it relates to software selection.

In summary, this research investigated a possible link between long-term retention and instructional technology for practical reasons. This was done so as to be able to give specific reasons to encourage or discourage the use of instructional technology in engineering (e.g. IT use results in increased long-term retention). Literature that describes projects with similar aims is presented in this chapter while literature that focuses more on the internal workings of memory and cognitive structures is not included.

2.1 Knowledge Retention

Knowledge retention is an essential part of education. As Semb and Ellis point out, "the very existence of school rests on the assumption that people learn something of what is taught and later remember some part of it," (1994, p. 253). This is particularly true in engineering education. In most engineering curricula, classes are built upon a foundation of information presented in other classes. In fact, most upper division engineering courses have multiple prerequisites, which, in turn, have prerequisites of their own. The fact that previous courses are required in order to take more advanced courses is founded on the assumption that students will retain, or remember, the information that was presented in the earlier classes. Yet, studies on retention have shown that this assumption does not always hold true. Furthermore, some studies have shown that knowledge retention can be improved through use of novel instructional methods such as educational technology. This section defines retention, quantitatively relates retention to traditional instruction, and presents some methods for improving retention.

2.1.1 Retention Defined

Knowledge retention is the recall or remembrance of information, processes, or skills that were once learned at a later point in time (Semb and Ellis, 1994). It is important to mention the distinction between retention and transfer. Retention is simply the ability to remember information as it has been presented, whereas transfer is the ability to remember information and apply that information to a new and distinct situation. Without adequate knowledge retention, transfer is nearly impossible. As such, the study presented herein focused solely upon retention and will pave the way for future studies investigating long-term knowledge transfer.

There is another type of retention that is of great importance to educators: the retention of students. Student retention is defined as the number of students who remain enrolled in a program or major. When students drop out of school or change majors, they are not retained. Student retention is also of special concern to engineering educators because many students transfer out of engineering due to the rigorous curricula. This type of retention was not addressed in this project. Therefore, within the context of this document, retention always refers to knowledge retention, not student retention.

2.1.2 Studies on Retention

A study into the literature on retention usually yields three main points. First, very few practical studies on retention have been successfully completed. Second, knowledge retention is often poorest when lectures are the primary source of instruction. Third, alternative instructional practices have been shown to improve retention to varying degrees. Each of these three points is discussed in more detail below.

2.1.2.1 Lack of Practical Retention Studies

At the beginning of this chapter, the point was made that there are two separate areas of research in knowledge retention, one being theoretical and the other being practical. Theoretical retention studies are common in the areas of psychology and cognitive science. These experiments usually take place in a laboratory and often test a subject's ability to recall simple information such as words, phrases, statements, simple relationships, symbols, etc. (see Ausubel, 1968, for numerous examples). These experiments often use very small retention times. For instance Moreno et al. (2001), who used various computerized agents to present botanical information to college students, tested the students' ability to remember botanical facts (i.e. retention) just minutes after the intervention. The reason for keeping retention times short and for focusing on simple information is because these experiments are conducted in the laboratory. In such laboratory settings, time and resources are limited. Participants are volunteers who receive some sort of incentive (e.g. money or course credit) for participating. The incentives are often limited and the participants can be very transient. As such, it is often desirable to complete the intervention as well as the assessment in a single session (Semb and Ellis, 1994). It is difficult to present large amounts of information or complicated information in single sessions. This is in contrast to natural or classroom research where retention intervals can be much longer and information can be presented over greater periods of time.

Although laboratory experiments are common, Semb and Ellis (1994) point out that validity is sacrificed for control. This means that while laboratory experiments in retention are often tightly controlled, they may not reflect the true and complex nature of education. In contrast, studies in natural settings (i.e. the classroom) are more difficult to control; it is much harder to control for non-experimental variables such as prior knowledge and ability. Whitley (1996) agrees that control and naturalism are often at odds and that choosing a particular type of research often results in a tradeoff. A result of this tradeoff is the dichotomous relationship between theoretical and functional retention studies. Critics of theoretical or laboratory experiments on retention suggest that the results of these experiments are not transferable to the classroom (Semb and Ellis, 1994).

Because the results of laboratory studies are no necessary valid in educational settings, practical retention studies must also be completed. Practical studies, however, are much less common than laboratory studies (Neisser and Hyman, 1999, Semb and Ellis, 1994).

Neisser and Hyman introduce their book *Memory Observed* by pointing out that the study of memory has very little to show for over a hundred years worth of research (1999). They suggest that the naturalistic study of memory may provide more applicable results than laboratory studies. However, naturalistic studies of retention in education are difficult and time-consuming (Hesketh, Farrell, and Slater, 2003) and thus somewhat sparse. Indeed, Neisser and Hyman state that “it is difficult to find even a single study, ancient or modern, of what is retained from academic instruction” (1999, p. 5). Semb and Ellis (1994), in response to an earlier version of Neisser’s book, state that the situation might not be as dire as he suggests but that they did have significant problems locating relevant articles on retention in academic settings. Neisser and Hyman (1999) state that this is not only because longitudinal studies of knowledge retention are complicated and time consuming but primarily results from the reluctance of psychologists to relinquish the amount of control that they have in laboratory settings for more natural research. Whatever the reason, there is agreement on the fact that there is still research to be done in this area. This is not to say that no research has been done (examples presented below point to the contrary), it is simply a call to researchers to conduct more practical, classroom experiments in order to determine the actual nature of retention in education (Neisser and Hyman, 1999).

This apparent lack of foundational research is not limited to naturalistic retention in general, but is also evident in the area of technology and its effects on retention.

Spellman (1999) points out that most research projects in the area of computer-assisted learning (CAL) include neither a measure of performance nor quantifiable data on the results of CAL. Lalley (1997) agrees and states that while the computer has become an important instructional tool and will continue to become even more important, there is little research in education to guide the implementation of educational technology.

In conclusion, research on the retention of knowledge taught in school is much less common than laboratory-based retention research and a need exists for such research to be conducted. Results from such research could have practical effects on the nature of education. Furthermore, research concerning the effects of technology on education is also rare but sought after. Examples of assessments conducted on educational technology presented later in this chapter will confirm these statements.

2.1.2.2 Knowledge Retention and Instruction

What little research has been conducted in the area of knowledge retention suggests that it is generally very poor. This may be linked to the fact that lectures remain the most popular teaching technique in higher education (McKeachie, 1999). This is an important point because research has shown that lecturing, when compared to other forms of instruction, result in the lowest levels of retention (Elshorbagy and Schonwetter, 2002). McKeachie (1999) agrees and adds that when knowledge is measured immediately following the educational experience, there is often no difference between lectures and alternative instructional techniques. When knowledge is measure some time after the experience, that is to say when retention is measured, lectured students usually perform worse than students who have received alternative instruction (McKeachie, 1999).

Quantitatively, Elshorbagy and Schonwetter (2002) state that students generally remember 70% of the first ten minutes of a lecture and only 10% of the last ten minutes. They further assert that ultimately only 5% of lecture material is retained (Elshorbagy and Schonwetter, 2002). Biggs presents the quantitative data in Table 2.1 and admits that while the numbers may not be hard and fast, they point out that listening to a lecture does result in smaller amounts of learning than other methods (1999).

Table 2.1 Amount of Learning vs. Instructional Method (source: Association for Supervision and Curriculum Development Guide, 1988 as cited in Biggs, 1999)

Most People Learn…	
10%	of what they read
20%	of what they hear
30%	of what they see
50%	of what they see and hear
70%	of what they talk over with others
80%	of what they use and do in real life
95%	of what they teach someone else

There are studies within engineering education that support these claims as well. A study performed by Bertz (1998) revealed that engineering students generally have very poor retention of elementary principles and low ability to transfer knowledge from previous courses. This is probably due to the fact that engineering educators, like their peers from other fields of study, rely heavily on traditional classroom techniques (i.e. lecturing) to present information.

There is evidence that the numbers presented above may be over-exaggerated. Semb and Ellis (1994) suggest that retention of information taught in school is not as poor as in laboratory settings which are usually the basis for figures such as those presented in Table 2.1. They do, however, acknowledge that forgetting does take place over time and that there are strategies and methods that instructors can employ to minimize the information loss. Some of these strategies will be presented in the following section.

2.1.2.3 Improving Knowledge Retention

There are studies in psychology, cognitive science, engineering education and elsewhere, that suggest methods for improving knowledge retention. Though suggestions vary in practicability and specificity, there are a few that are generally accepted among educators. Some of these accepted methods are presented and discussed in this section.

Activity

As Biggs (1999) anecdotally points out in Table 2.1, in order for students to learn and retain more than 50% of the educational material, they must do something. Furthermore, Biggs says that "being active while learning is better than being inactive" and that "activity is a good in itself," (1999, p. 76). Semb and Ellis reached a similar finding and stated that instructional "strategies that more *actively* involved students in the learning process" yielded increased amounts of differential retention (1994, p. 277, emphasis added). While the term *active learning* refers to a specific school of thought and research in education (as presented in Kenimer and Morgan, 2003 and Felder and Brent, 2003), the type of activity spoken of by Biggs (1999) is more general and includes strategies as simple as holding in-class discussion groups. Semb and Ellis (1994) found that in their research, any type of activity that produced a qualitative difference in the learning experience resulted in greater retention.

Other research into learning and retention has also shown the need for activity. Dale (1968) revealed that there are three different levels of experience and learning: enactive, iconic and symbolic. Enactive learning involves direct experience, actually doing something such as tying a knot. Iconic experiences are those that involve pictures or graphics, such as diagrammatic instructions on how to tie a knot. Symbolic experiences are those that include abstract symbols (typically words and languages), such as the word *knot*. Dale (1968) expands on the discussion of activity by pointing out that experiential learning is not only more rich, but is a prerequisite for more abstract learning; the word knot, for example, has more meaning when one has encountered a picture of a knot and is even more meaningful if the learner has actually tied a knot.

These three levels of experience relate directly to Tulving's (as cited in Biggs, 1999) three memory systems: the procedural memory where actions are learned and retained, the episodic memory where images are learned and retained, and the semantic memory where declarative knowledge is learned and retained. Biggs (1999) points out that these three systems do not operate in identical manners and that data stored in procedural memory is the easiest to recall and data stored in semantic memory is the hardest to remember. Synthesizing Biggs' and Dale's research reveals that combining instruction with activity not only increases the amount of the instruction that is retained but also ties the activity with other abstract instruction, thus increasing retention of related material.

On a final note, Biggs makes two important notes about introducing activity. The first is that any type of activity could be beneficial because it breaks up monotonous lectures and revitalizes students' attention spans (Biggs, 1999). The second point is that the benefit that is realized by introducing activity is vastly increased when the activity is directly related to the instruction because creates a link between the three memory systems (Biggs, 1999). In summary, research has shown that using any type of in-class activity increases retention and if the activity is well designed and related to the instruction, retention is increased to an even greater degree.

Advance Organizers

David Ausubel, a cognitive scientist who has spent decades studying learning and knowledge retention, stated that retention is improved when learning is connected to what has previously been learned (Ausubel, 1968), a finding that Semb and Ellis (1994) realized in their research as well. One way to accomplish this is through the use of advance organizers. An advance organizer is a cognitive tool that links what is already known to what is going to be subsequently taught (Ausubel, 2000). Careful planning and organization are required for proper application of advance organizers. An understanding of what students already know is critical to the process of connecting old material to new information. The function of the organizer is not to simply introduce new material or to review what was taught in previous lectures In addition to serving these two purposes, advance organizers provide direct relationships between the two so that the new material builds upon and is connected to the existing cognitive structure. Ausubel (2000) provides further details about what constitutes an advance organizer and guidelines for applying them properly, though an extensive discussion on the topic is outside the scope of this research. When used correctly, research has shown that advance organizers have a positive effect on learning (Ruthkosky & Dwyer, 1996, Ausubel, 2000). Ausubel admits, however, that studies investigating the long-term outcomes stemming from the use of advance organizers are sparse.

Alternative Instruction

Smith (as cited in Issa & Domitrovic, 1995) showed that the way in which information is disseminated can have an impact on knowledge retention. As mentioned above, traditional instruction (i.e. lecturing) has been found to be the least effective method of teaching when retention is measured. A number of different alternative teaching methods have been studied for their impacts on retention. Son and VanSickle (2000), for example, found that using a problem solving approach to teaching improved both performance and knowledge retention four weeks after the instruction. The problem solving model Son and VanSickle developed involved teaching domain-specific knowledge within the context of a well-formulated, complex, real-world problem (2000). Similarly, Silverstein and Baker taught calculus concepts to engineers in the context of engineering problems with the intent of improving retention (2003).

Elshorbagy and Schonwetter (2002) recommend using inductive instruction in engineering education. Inductive instruction, which is based on constructivist theory, is also referred to by Elshorbagy and Schonwetter as reverse lecture (2002). This is because traditional lectures begin with abstract theories or principles from which specific applications, equations, or examples are derived for practical use. Inductive instruction works in reverse, the instructor presents specific examples or applications to students and then facilitates the students in developing abstract theories from these examples. Elshorsbagy and Schonwetter state that this form of instruction can have a great impact on students' retention of knowledge. Hesketh et al (2003) also suggest that a carefully planned inductive approach to teaching can increase retention, though no specific evidence is presented.

Practice

Gattis found a strong correlation between supplemental instruction attendance and knowledge retention (2000). Supplemental instruction (SI) is a program where instructional sessions are scheduled in addition to regular classes. Students attend these sessions on a voluntary basis and are given additional instruction on course topics by tutors or teaching assistants (TAs). Gattis was unable to determine what exactly was the cause of the increased retention though he was able to rule out motivation (i.e. he was concerned that only motivated students attended SI and thus had greater retention rates but was able to reject this possibility). He did, however, postulate that guided practice, which is defined as studying course topics under the guidance of a more knowledgeable person such as a TA, was a significant contributing factor (Gattis, 2000). It seems intuitive that as students spend more time learning and solving problems within a particular context that their retention of knowledge within this context would increase. The work of Gattis supports this conclusion.

Instructional Technology

Many studies have been performed that compare the effects of instructional technology to those of traditional lecture, though few continue those comparisons at a later date to determine the effects on retention (please see section 2.3 below for more on this). One example of a study that included retention was performed by Yildirim, Ozden, and Aksu (2001) who compared the use of hypermedia to traditional lecture. It is theorized that a person's cognitive structure is similar in organization to hypermedia programs. Thus, Yildirim et al. further hypothesized that learning through the use of hypermedia would increase retention. They compared students who learned biology material through the use of a hypermedia learning environment to a comparison group who learned the same material through typical classroom instruction. There was no difference in performance immediately following the instruction, but the hypermedia group did perform significantly greater on a retention test, one month after the instruction, of declarative, procedural, and conditional knowledge.

Studies within the context of engineering have also proposed a link between instructional technology and higher retention. Riley and Pace (1996) suggest that the use of technology in engineering education "often [improves] the effectiveness and efficiency of instruction" (p. 366). Others, including Sulbaran and Baker (2000) and Issa and Domitrovic (1995), also hypothesize that using instructional technology can improve students' retention of engineering information. This is generally assumed because it incorporates principles of activity, alternative instruction, and practice as well as advanced visualization and simulation techniques (Hmelo, Lunken, Gramoll, and Yusuf, 1995). As these principles alone have been shown to increase retention, it is assumed that well designed and appropriately applied instructional technology could improve retention as well. As will be shown below, however, more evidence is needed to support these hypotheses.

2.1.3 Summary

Three major points arise from the literature on knowledge retention. The first is that most studies of retention involve laboratory experiments, which have been shown to exaggerate the amount of information that is lost over time. Less common are naturalistic studies conducted in the classroom, which are not only more applicable to educational settings but also more accurate concerning information loss. The second point is that traditional, lecture based instruction yield lower levels of retention than alternative methods. The third and final point is that there are strategies that can improve retention including activity, advanced organizers, alternative instruction, practice, and possibly instructional technology.

Investigating instructional technology and its impacts on long-term retention became the objective of this study. Before the development of this objective is presented, however, a short review of literature on instructional technology in engineering education is presented. Included in this review are examples of IT and any assessments that have been conducted on these examples to determine their impact on learning and retention.

2.2 Instructional Technology

Instructional technology is the broad term used to describe any type of computer-based technology used in an official capacity for instructional purposes in courses. In the literature, the terms multimedia and instructional technology are often used interchangeably, though multimedia refers to a broader range of technologies (e.g. television) and not all instructional technologies take advantage of different kinds of media, some are purely textual in nature. Where the term multimedia is used in this document, it is computer-based multimedia that is being referred to and can thus be considered instructional technology.

There are, of course, exceptions to this definition of instructional technology. The first exception includes situations where computer use is the end goal and not an educational enhancement, such situations are not considered IT (for example AutoCAD would not be considered IT in a course with a goal to teach computer aided design). Another exception is the use of computers to perform routine and required calculations. Based on these two exceptions and on Semb and Ellis' research, IT is more broadly defined as any use of computer-based technology that results in an educational experience that is qualitatively different than traditional, lecture-based instruction. Under this revised definition, examples of IT include such uses as PowerPoint presentation slides during lectures, structural analysis software used to complete course projects, and intelligent tutors to help students understand course topics. These three examples, however, would fall into different categories of IT as explained below.

2.2.1 Types of Software Used in Education

Glennan and Melmed (2000) classify educational or instructional technology into three categories or types of software: tools, content, and instructional management. Tools are applications or packages that have been developed for purposes other than education. They have specific tasks and are usually found in commercial or home settings. Examples of tools are word processors, spreadsheets, and structural analysis software.

The second type of IT is content-type software. Content-type software are packages that have been developed specifically for instructional purposes. These packages are developed for use in educational settings and can include intelligent tutoring systems, online simulation materials, or electronic textbooks.

Instructional management software packages are designed to assist in the administrative duties of a teacher, such as relating coursework to curricular requirements, tracking student progress, and maintaining course calendars. Because these tools are not used to convey any type of material to the students, instructional management tools were not included in this study.

2.2.2 Examples of Instructional Technology in Engineering Education

This section provides a number of examples of IT use in engineering education. While these examples are not exhaustive, they are exemplary. The examples below do represent typical situations and implementations of IT in engineering education. The examples are categorized by being either content-type software or tool-type software as defined in the preceding paragraphs.

2.2.2.1 Content-type Technologies

Content software is continually being developed and encouraged for use in engineering education. Some of the examples are given here. Assessments of these implementations will be presented later in this chapter.

Riley and Pace (1997) implemented multimedia to improve the efficiency of classroom instruction and to present complex concepts with animations and photos. Their implementation involved developing PowerPoint presentations for all lecture materials. These presentations were used during class and were also posted on a server so that students could access them on their own time.

Issa, Cox and Killingsworth (1999) used an interactive CD-ROM to teach construction safety to both undergraduate and high-school students. The software had different modules to present information on various areas of safety and utilized video and other presentation media. The software was also interactive in that students could control the pace of the software and were required to answer questions throughout the training process.

Samek and Landry (1997) used computer-based simulations in mechanics classes to illustrate complex concepts. Specifically, simulations were created using Working Model™ and Mathcad™ to demonstrate topics in dynamics including instantaneous center of zero velocity and zero velocity condition at the contact point of a non-slipping wheel. These are topics which they had found their students struggling with and so they developed models with animated dynamic systems to show how velocities varied in the system and how these velocities changed over time. Their animations were not interactive but were developed to help students visualize the concepts.

Many instructors have used the World Wide Web to post lecture notes or provide students with visual information. Dymond (1996), for example, maintained a web site of course information divided up into core concepts with links to external sites that contain additional content pertinent to the topic or concept. As described previously, Riley and Pace (1996) posted the PowerPoint lecture notes on the web so that students could go back and review them at their own pace.

Another type of computer-based instructional tool created by Aminmansour (1996) was an electric textbook with distinct chapters and exercises for students to work on. The software was developed for use in a steel design course, was referred to by the developers as intelligent courseware, and was designed as supplement, not a replacement, to standard course texts and lectures (Aminmansour, 1996). The software utilized full motion video, graphics, animation, and audio to present information. This information was categorized into chapters, like a textbook, and was interactive in that is asked users questions at various points throughout presentation.

Thukral and Gramoll (2001) set up an online Engineering Media Lab complete with electronic textbook, shockwave animated lectures, java applet calculators (to calculate unit vectors, cross products, truss member loads etc.), quizzes and tests. In addition to all these capabilities, it also included an instructional management tool which allowed the instructors to track student grades and communicate with the students easily. Communication and collaboration tools were also built in which allowed students and instructors to post messages on bulletin boards or communicate synchronously in chat rooms complete with drawing and calculating tools. The Engineering Media Lab currently supports content for Statics and Dynamics classes taught at four different universities, with more content and subscribers scheduled for future semesters.

Sulbaran and Baker (2000) created a virtual construction environment that allows users to visualize construction situations in three dimensions and from many different angles. Specifically, the environment was designed to help further student understanding of the crane selection process. The environment was posted online and used in graduate and undergraduate construction management classes as a visualization tool to help students see cranes in a more natural setting.

Finally, one other package presented actual engineering projects to students in the form of multimedia-supported case studies (Angelides, Poulopoulos, Avgeris, & Haralampous, 2000). The case study included complete project information from an actual engineering project with analysis and design information being supplied by the developer. The case study was put online for students to use in a senior level course where aspects of planning, design and construction are taught and was designed to help students relate their educational experiences to real life problems (Agelides et al., 2000).

2.2.2.2 Tool-type Technologies

Tool-type technologies are tools that were developed for non-educational purposes but may be used in classroom settings as well. These tools generally have a 'real world' application. Some examples of tool-type software used in engineering courses are presented in this section.

Hein and Miller (1995) had students use the programming and presentation aspects of spreadsheets to enhance their structures courses. They developed learning modules in a common (unnamed) spreadsheet program that contained animations, graphics, simulations, calculations, audio clips, charts, and diagrams for various topics within structural mechanics. The modules were used in class in the hopes that students would spend more time studying the behavior of structures and less time copying notes from the board. It was further envisioned that these modules would allow students to interact with more structures than in a traditional lecture.

More common is the use of actual engineering design software in the classroom. Meyer and Ressler (1995) used two structural design packages in a steel design course. Specifically CME-Truss was used for complex homework projects involving the design of a truss for maximum economy. Students used the software to design trusses and the instructor, who had an add-on loading module, would load the trusses in class to determine the truss capacities. Another program, LRFD92, was used for the design of steel beams, columns, and beam-columns. Students used LRFD92 to verify hand results for design projects.

One unusual tool (Muscarella, O'Neill, & Gano, 1997) not only performed complex structural matrix operations for the students but also presented how and why the operations are being done, thus instructing and calculating simultaneously. Students used the software by inputting member local stiffness matrices and the STACKER program showed the students step-by-step how these were transformed into the structural stiffness matrix. The overall matrix could then be exported for use in a program capable of performing matrix algebra to obtain various solutions to structural problems.

Dr. Frame (Dr. Software, 2001), a 2-D structural modeling environment that is used in a number of different industrial settings, has also been used extensively in many national and international universities to enhance structures courses. Dr. Frame, and its companion piece of software Dr. Beam, were used to analyze two dimensional structures and elements including beams, columns, trusses, and frames made from a steel (using standard AISC shapes), concrete or other materials. Dr. Frame has been used in classrooms to both verify hand calculations, and to allow students to solve a large variety of problems in the same amount of time required for a single hand solution, thus exposing students to numerous structural systems and designs.

As these are just some examples of IT in engineering, both content- and tool-type; this list is surely not complete. Many other software titles are used in many other instructional capacities both in engineering and in other fields of study as well. Of the implemented software titles exampled in this chapter, some were assessed to varying to degrees to determine their impact on learning and retention. These assessments and their results will be summarized to show that the link between IT and long-term retention, as hypothesized in Section 2.1.2.3 of this document, has yet to be established.

2.3 Instructional Technology and Retention

The examples mentioned above represent just a few of the ways technology is being implemented in engineering classrooms. Few of the researchers involved in the examples, however, conducted assessments to determine the effectiveness of the IT that was implemented. This section will revisit each of the preceding examples to determine the degree to which the intervention was assessed as well as the results of those assessments. As before, this section is divided by technology type, with an addition section added on IT used in fields other than engineering.

2.3.1 Assessment of Content-type Software

The assessments of the content-type technology examples listed in Section 2.2.2.1 were completed at different points in time and under varying degrees of control. Riley and Pace (1997) assessed their use of multimedia slides using the following design. Roughly 45 students (as with most classroom research the numbers fluctuated based on attendance) were divided into two groups designed to be academically equivalent based on previous course grades. A case study in concrete construction was presented to the students in one group via static overhead slides while the other group was presented with an animated version of the process. Students were assessed immediately following the presentation to determine how many of the steps they recalled from the construction process. Students were then assessed for retention via an unannounced quiz of factual information during the following lecture period. The researchers reveal that the multimedia group had higher average scores on the posttest and the retention test and thus declared the multimedia use a success. There were, however, no statistical tests conducted to determine if these differences were due to chance or an actual result of the intervention. In fact, the multimedia group outperformed the lecture group by only 5/100 of a point, or 1% of the total points possible—a difference that most likely would not be significantly different. Because the results of the assessments were not analyzed in a scientific and meaningful manner, no conclusions can be drawn from the data. Unfortunately, though an assessment was completed, it had very little efficacy.

Issa, Cox, and Killingsworth's assessments were more meaningful (1999). In their study, 52 students were broken into smaller groups of about 10 students who would learn construction safety material by either attending a classroom lecture or by interacting with a CD-ROM in the computer lab. As there were a number of different safety topics, students who attended the computerized version of one topic would then attend the class version of the following topic and vice versa. Students were assessed with a posttest immediately following the intervention and then a retention test three weeks later. Statistical analysis of the results revealed that students in the computerized sessions performed significantly better on both the retention test and the posttest.

Sulbaran (2002) assessed his use of virtual reality in the classroom in a rigorous manner as well. He randomly assigned 70 students to three different groups who each received crane selection instruction via one of three different methods: traditional class lectures, web-based materials, and virtual reality materials. Retention was assessed through the use of a posttest administered in the session of class immediately following the intervention. The results revealed no significant differences on posttest scores between the three groups.

While these three studies did include an assessment, none of them included long-term assessments. If students are expected to remember information until at least the end of the semester and ideally well into the next semester, long-term assessments with similar time frames should be included to determine such effects. Most of the content-type examples mentioned, however, included no assessments at all. Samek and Landry's paper (1997) on the use of working model simulations in class did not include any assessments or any mention of future assessments. Dymond's discussion (1996) of using the Internet for the posting of course material also did not include a discussion on assessement. Animansour included no mention of assessments or evaluations in his paper on the use of e-textbooks in the classroom either (1996). Angelides et al. (2000) similarly focused on the implementation of the technology, specifically multimedia case studies, and made no mention of evaluating the effectiveness of the technology. Finally Thurkall and Gramoll (2002) also focused on how the technology could be used but did not assess the effectiveness of that technology.

2.3.2 Assessment of Tool-type Software

As with the content-type software, the assessments of tool-type technologies listed in Section 2.2.2.2 were completed to varying degrees. A few different Dr. Software modules were assessed during the developmental stages by Miller and Cooper (1995). The assessment was conducted by comparing final exams completed by students in a traditional, lecture-based course to students in a course that routinely used Dr. Software modules in addition to alternative instruction techniques. The researchers, however, admitted that the assessments were flawed for a few reasons. One reason was that different instructors taught the different courses, one who was experienced in teaching the course and one who was teaching it for the first time. Another was that the final exam was worth a different percent of the students final grades in each of the two courses, 16% in one and 40% in the other, which could have affected student motivation. There were other flaws as well but perhaps the most significant is the lack of control. Because alternative instruction and various software titles were used in the class, there was no way to separate the effects of the different variables. As such, the results of this assessment were not meaningful.

Muscarella et al. (1997) used student perceptions as the sole form of assessment for their STACKER program. After using the technology, students were asked to fill out a questionnaire about their abilities and opinions of the software. Actual achievement resulting from the software use was not measured and thus no meaningful results about learning and/or retention were obtained.

The other two tool-type examples, Hein and Miller (1995) and Meyer and Ressler (1995), did not include any mention of assessments of learning. Both of these examples were focused on showing how tools could be used in education. Neither, however, evaluated whether such use was actually beneficial.

2.3.3 Discussion of Assessments

Of the programs listed in the previous two sections, only two performed any meaningful assessment of IT use. These two results were at odds; Sulbaran (2002) found no difference in pretest scores for IT groups, but Issa et al (1999) found that students who used IT performed better on posttests and three-week retention tests. The reasons for these discrepancies are unknown. A deeper look into each of the studies may reveal the nature of the differences, but such an analysis is outside the scope of this work.

There are two major conclusions that were drawn from this investigation into the assessment of IT use in engineering. First, most of the studies included assessments that were not meaningful or no assessments at all. The second is that those that did include assessments, did not include assessments on long-term retention. The longest retention interval was three weeks, a minimal amount of time compared to how long instructors generally expected their students to retain course information. Retention of information well into subsequent semesters and beyond is a goal of most instructors and it should be a goal, and thus a measure, of any educational intervention as well.

Most of the examples listed above focused on use and implementation rather than effectiveness of use. This phenomenon is not isolated to the examples listed. Bouchlaghem, Sher and Beacham (2000) cited five instructional programs that they considered to be successful. In their article, they encouraged the use of these software technologies in courses. Unfortunately, only one of the programs was formally assessed for learning, but not for retention. It should cause concern that software programs are being considered successful and useable and being encouraged for use without any testing done to determine the impact that they have on education. Semb and Ellis (1994) strongly recommend that any novel instructional technique should be assessed not only for its impacts on learning, but on long-term retention as well. In engineering education, especially in the area of IT, assessments of learning impacts are rare. Furthermore, this literature revealed no studies in engineering education that investigated the long-term impacts of IT.

2.3.4 Long-term Assessments in Other Fields

The lack of long-term retention studies in IT, however, is not limited to the field of engineering. Spellman (2000) evaluated the use of computer aided learning in college level Geography courses and while he stated that most IT projects "have not included a mechanism for performance measurement" which has resulted in little quantifiable data, his project only measured attitudes as well; performance and retention were not assessed. Ubuz (2001) used an interactive computer program to teach college students three key calculus concepts. Ubuz did measure performance at the end of the computer-aided intervention but a follow-up retention test was not administered. Moshell and Hughes developed an Internet-based multimedia domain for experimental learning and conducted a few experiments with elementary school children "but no formal evaluations of educational effectiveness have been performed" (1996, p. 104). Daily (1994) used multimedia to enhance courses in management. Daily tested students throughout the courses, thus assessing performance, but did not test retention after the intervention was complete.

There are some studies that do include retention assessments, though with retention intervals of no more than four weeks. One study used a computer tutorial to teach soldiers how to build a specific type of radio (Orey, Zhao, Fan, & Keenan, 1998). The study compared those who had used the tutor to those who had learned by hands on experience and gave the subjects a surprise quiz four weeks after the intervention to see if there was a difference. Both the posttest and the retention test involved the actual building of the radio while experts rated and scored performance. The subjects who used the tutor outperformed those that had received hands on experience on both the posttest and the retention test.

Similarly, Durham and Emurian (1998) used a retention measure after four weeks to compare subjects who had used a command line interface to subjects who had used a menu interface when performing programming tasks. Retention was measured by counting the number of requests for help, counting the number of errors, and measuring the time taken to complete the task. Both groups had forgotten some information but the command line group retained significantly more than the subjects who had used the menu-driven interface did.

Williams and Zahed (1996) also tested for retention after a month when comparing subjects who received safety training via a computer tutor with others who received traditional training. Performance and retention were measured through the use identical, multiple-choice tests. In this study, the computer group performed significantly better on the retention test than did lecture group. The groups performed equally well on the posttest.

One-month retention tests have also been used to compare hypermedia instruction to traditional instruction in a ninth grade biology class (Yildirim, Ozden, & Aksu, 2001). In this study, there were no significant differences in posttest scores between the lecture and hypermedia group. The hypermedia group did, however, perform significantly better on the retention test one month later.

Lalley (1998), however, tested retention after just one week when comparing the differences between text and video feedback in computer tutorials aimed at students in a middle school science course. A multiple-choice posttest was administered immediately following the intervention and four weeks after the intervention. Students who received video feedback during a computer tutorial performed significantly better than those who received textual feedback on both the posttest and the retention test.

2.3.5 Discussion of Assessments

As with the engineering examples, some of the non-engineering examples did assess for learning and retention outcomes while others included no assessments at all. This lack of retention studies reveals that, similar to the engineering studies, technology is being implemented in fields other than engineering without any assessments being done to determine the educational effectiveness of these implementations. Also like the engineering examples, the studies in other fields had relatively short retention intervals, with the longest being four weeks. Again, this is shorter than the amount of time that most instructors would hope for their students to retain course information.

The non-engineering studies that included retention measures were unlike their engineering counterparts in one important way, the setting in which the experiments took place. Two of the studies, Lalley (1998) and Durham and Emurian (1998) were conducted in laboratory settings, which according to Biggs (1999) and Semb and Ellis (1994) yield fundamentally different results than studies conducted in naturalistic settings. Two of the remaining three studies, Orey et al. (1998) and Williams and Zahed (1996), were in naturalistic settings, but these were training settings as opposed to educational settings, which differ in a number of ways as well. Finally, the last retention study, Yildirim et al. (2001) was conducted in ninth-grade classes, which are also fundamentally different settings than college classrooms.

2.3.6 Summary and Conclusions

The examples presented in this section were given to show that there remain gaps in the theorized link between IT and knowledge retention. Much research in IT, both in engineering and in other fields, do not include measures of retention. Some of the studies that do include measures of retention take place in fundamentally different settings than college-level engineering classrooms and thus may not transfer appropriately. Of the remaining studies, that do have retention measures and are in settings similar to college-level engineering courses, the maximum retention intervals are three to four weeks, an interval that is a fraction of a semester and shorter than desirable for most instructors. These gaps in the literature, combined with the information presented in Sections 2.1 and 2.2 lead to the development of the research objectives, which is the topic of the next section.

2.4 Development of Research Objectives

Based on the literature presented in the preceding sections, the research objectives that drove this study were developed. Following is a bulleted list of the main conclusions from the literature review. The points in bold represent gaps in the literature based on the main conclusions. It is these gaps that this research project was designed to fill.

- Laboratory experiments in retention are quite common
- Naturalistic (i.e. classroom) experiments in knowledge retention in education are quite uncommon
- Laboratory experiments in retention have been shown to not be valid in educational settings
    - As such, **more experiments in knowledge retention, especially with long-term intervals, need to be conducted in the classroom**
- Lecturing has been shown to be an inferior instruction technique when retention is measured

- A number of methods have been shown to improve retention, including activity, alternative instruction strategies, advanced visualization techniques, and practice
- Instructional technology has been theorized to improve retention because it includes elements of activity, practice, visualization, and alternative instruction
    - **The theorized link between instructional technology and knowledge-retention was not firmly established** in this literature review
- Many different forms of IT have been implemented in engineering and other fields of study
- Researchers who implemented IT rarely measured the effects of the IT on retention under naturalistic conditions
- Research that has been conducted on the effects of IT on knowledge retention usually have retention intervals of no more than 3-4 weeks
    - This literature review revealed **no classroom studies that assessed the effects of IT on retention at intervals that more closely resemble semesters in length of time**
- There are two different types of IT that students use, tool-type and content-type software
- Studies have investigated the use of tool-type software and studies have investigated the use of content-type software

- This literature review revealed **no studies that compared the effects of tool-type use to content-type use**

This research project was designed with the objective of filling the gaps in the literature, listed above in bold. The objective of this study was to design and implement a research project that would take place in the classroom and establish a link between IT and long-term retention by implementing IT in the classroom and then evaluating the students for learning as well as retention at 10 and 25 weeks.

Researchers in the fields of retention and instructional technology supported this objective. Lenox, O'Neill, and Dennis have stated that computer use had "permeated the math, science, and engineering courses in the civil engineering curriculum," (1995, p. 240). Simply using such tools, however, does not guarantee increases in student performance and retention (Krone, 1995). Lalley (1997) pointed out that there is little research in education to guide the implementation of such widespread computer use. Semb and Ellis agree, and more pointedly stated that "recent innovations in approaches to teaching and the application of instructional strategies should incorporate measures of both *learning* and *long-term retention* in the evaluation process," (1994, p.278, emphasis in the original). The implementation of instructional technology as exampled in this chapter and as completed in this research is considered an innovative approach to teaching and thus, as Semb and Ellis state, should be evaluated for learning and long-term retention. Furthermore, Semb and Ellis (1994) as well as Biggs (1999) recommend that this research be done in classrooms, the setting to which the results will later be applied.

Although Semb and Ellis do not define the length of the retention interval required to constitute long-term retention, some conclusions can be made based on their work. First, they emphasize the performance of studies under naturalistic conditions. With this in mind, it was concluded that the long-term retention interval should be naturalistic as well. Based on this assumption, a retention interval of 10 weeks was chosen because this represents the interval between the designed intervention and the end of the semester. A long-term retention interval of 25 weeks was chosen because this represents the interval between the designed intervention and a point in time during the subsequent semester when the information may be needed again for application to new and more complex topics. These intervals are consistent with the projects in Semb and Ellis' analysis (1994), which included a number of retention studies—none of which studied IT, however—with intervals of 20-50 weeks. This is further evidence that Semb and Ellis, when calling for long-term retention studies are calling for retention intervals similar to those used in this research.

One other objective that stemmed from the gaps in the literature was to design and implement a research project that would investigate the different retention effects that could result from implementing a tool-type technology as opposed to a content-type technology. This was an objective that was developed because both types of software are being implemented in the classroom, but as Lalley (1997) states there is no educational research to guide this implementation. The results of a study investigating these differences could serve as such a guide to implementation.

These two objectives translated into the two main research questions posed at the beginning of this chapter and restated below. In the following chapter these research questions will be developed into formal hypotheses. The development of the research designed to answer these questions is also presented in the following chapter.

1. Does the use of instructional technology in the engineering classroom increase long-term knowledge retention?
2. Is there a difference in the long-term effects when using a tool-type software as opposed to a content-type software?

CHAPTER 3

PROPOSED METHODOLOGY

The methodology implemented in this research project differed from the proposed methodology in a number of ways. The reasons for the changes in methodology were mainly due to concerns raised by instructors of the courses in which the intervention took place. This chapter discussed the methodology as it was proposed to and approved by the guidance committee. Chapter 6 will discuss how the project was actually implemented and why certain changes were made.

3.1 Research Questions

There are two questions that this research project was designed to answer. First, do students learn and retain more information when using instructional technology as compared to traditional classroom techniques? Second, do students learn and retain more information when using a content-type software as compared to using a tool-type software?

3.2 Variables

Three types of variables are discussed in this section: dependent variables, independent variables, and non-experimental variables.

3.2.1 Dependent Variables

This study researched the impact of technology on three different variables: learning, retention and long-term retention. Learning (as a variable) was substantively defined here as the amount of knowledge obtained during the intervention. Learning was originally defined operationally as the difference between performance (i.e. score) on a posttest that took place immediately after the intervention and performance on a pretest that took place prior to the intervention. Performance and learning are used interchangeably throughout this document to represent the same variable.

In the context of this work, retention was substantively defined as the amount of information retained after a period of ten weeks. Operationally, it was defined as the difference between scores on a retention posttest, taken ten weeks after the intervention, and the aforementioned posttest, taken immediately after the intervention. Long-term retention was substantively defined as the amount of information retained after a period of 25 weeks and was operationally defined as the difference between scores on a posttest, taken approximately 25 weeks after the intervention, and the posttest that followed the intervention.

Thus, four tests were actually administered: a pretest, a posttest, a retention posttest, and long-term retention posttest. The dependant variables were proposed as differential scores. Figure 3.1 illustrates how these four tests and three variables are related.

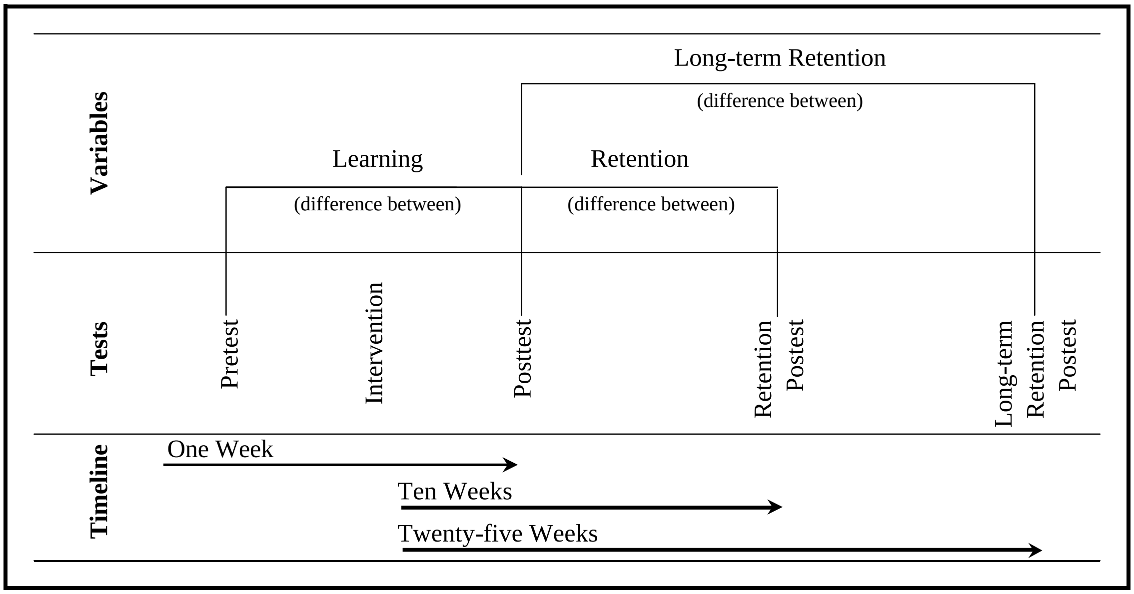


Figure 3.1 Measurement Variables

3.2.2 Independent Variables

Again, the objective of this research was to investigate a possible relationship between the use of technology in engineering education and knowledge retention. As such, the independent variable was the use of technology. More specifically, this work sought to compare the use of tool-type technologies to the use of content-type technologies (as defined in Chapter 2) in the engineering classroom. Thus, there were three independent variables or experimental conditions: 1). the use of a tool-type software, 2). the use of a content-type software, and 3). no software use at all. The procedure section below will discuss the experimental conditions in more detail.

3.2.3 Non-experimental Variables

In order to establish a link between instructional technology use and learning and retention, a number of other variables were controlled or accounted for. Although research involving human subjects, however, can never be completely controlled, a number of methods were followed in this research to avoid confounding influences from non-experimental variables. Some non-experimental variables of concern are briefly introduced here and are discussed in more detail in various other sections of this document.

- Prior contextual knowledge. It was anticipated that some of the subjects may have been familiar with the material covered during the intervention. A pretest of contextual knowledge was given to each of the experimental groups and the results were analyzed to ensure that none of the groups was biased by having more students familiar with the material than the other groups. The pretest and the results will be discussed in greater detail later in this document.
- Student ability. Some students generally perform better in educational settings than others. Having more exceptional students in one group than in the others would bias that group and confound the experiment. To avoid this, grade point averages (GPA) for each student were collected and compared across groups to ensure that the groups were similar.
- Instructor. It was assumed that not all of sections during which the intervention would take place would be taught by the same instructor. As such, a possible instructor bias could potentially be introduced. This was avoided by having one instructor teach the course topic that the intervention was designed for to all the sections involved. This will be discussed in more detail later in this chapter.
- Major and rank. Students from a number of different majors and schools participated in the intervention. In order to avoid a possible bias due to differing majors, the groups were compared to ensure that the majors were equally

represented in each of the groups. Also, students from all ranks were in each of the groups and they were similarly compared.

- Demographics. Information concerning gender and race were gathered for comparison purposes, however there is no reason to assume that either would have an effect on performance and retention and as such these variables were not of substantial concern.

3.3 Hypotheses

With the research questions and variables explicitly defined, the hypotheses can be formally presented. As stated in the previous chapter, research suggests that instructional technology may improve knowledge and short-term retention when compared to traditional lectures. This work was designed to add to that body of research as well as expand it to include long-term retention. As such, three hypotheses were as follows.

- Use of technology, whether tool or content type, in an engineering class setting will result in an increase in student performance.
- Use of technology, whether tool or content type, in an engineering class setting will result in an increase in knowledge retention.
- Use of technology, whether tool or content type, in an engineering class setting will result in an increase in long-term knowledge retention.

This research was also designed to compare the use of tool-type software to the use of content-type software. There is very little research, however, in this area on which to base an hypothesis. As such, the experimental hypothesis is that there is no difference between the two, although the actual purpose of testing this hypothesis was to determine whether a difference really did appear. As such, the final three hypotheses were:

- Use of a tool-type technology will not result in increased student performance when compared to use of a content-type technology when used in an engineering class setting.
- Use of a tool-type technology will not result in increased knowledge retention when compared to use of a content-type technology when used in an engineering class setting.
- Use of a tool-type technology will not result in increased long-term retention when compared to use of a content-type technology when used in an engineering class setting.

3.4 Domain

Careful consideration was put into choosing the domain in which the experiment would take place. The topic that was chosen as the focus of the intervention was the truss analysis portion of a statics course. The reasons for this decision will be discussed in this section.

3.4.1 Course Selection

Statics was chosen as the course in which to intervene for a few reasons. First, working with students in the statics courses allowed access to a large sample population. At Georgia Tech, where the experiment took place, an average of nine sections of a combined statics and dynamics course (CEE 2020) are taught during fall and spring semesters. Each section usually contains an average of 40 students though some sections have as many as 75 students enrolled. With the exception engineering graphics, no other course in the School of Civil and Environmental Engineering (the school that supported this research) contain as many students as CEE 2020.

Another reason for choosing CEE 2020 was due to the position of the course in the curriculum. Statics is often taken during the sophomore year. This is important because in order to assess long-term retention, the population had to be available for future studies. Picking a course that students took relatively early in their curriculum was necessary to prevent participant mortality and it facilitated in tracking students for future assessments because many of them were still on campus.

The third reason for selecting statics was universality of the course. Statics is the foundation upon which much of the Civil, Mechanical, and Aerospace Engineering curriculums are based and as such is required of students in these fields. It is also often a required, but not fundamental, course for other majors as well, including industrial, electrical, and textile engineering. The result of statics being a requirement for most students in engineering programs worldwide is twofold. First, a number of educational technology programs have been developed for use in statics. Because so many students worldwide take statics, a number of textbooks and software titles have been developed to cater to this large consumer group. Such a condition was necessary for this research because the focus of this work was assessment, not development, of educational technology. Having a number of software titles to choose from allowed the researchers to choose one that met the goals of the research without having to spend time in developing new software. More on software and selection thereof is discussed in the next chapter.

The second result of the universal nature of statics and the final reason for choosing it as the environment for this experiment is that the results can be generalized to a large population. With readily available software being developed for this course and thousands of students taking the course worldwide at any given time, the results of this research could clearly have an impact on many people.

3.4.2 Topic Selection

Within the broad domain of statics, the more specific topic of truss analysis was chosen as the focus for this experiment. The reason for focusing on one particular topic was to avoid the problems that prevented Felder, Felder, and Dietz (1997) from drawing any conclusions from their study on retention and innovative instruction as described below.

Felder et al. (1997) attempted to perform a longitudinal study of the effects of non-traditional classroom practices on knowledge retention. Alternative teaching approaches were introduced in an introductory engineering class at the sophomore level and the researchers had intended on tracking the students' performance throughout their continuing undergraduate studies (Felder et al., 1997). The researchers ran into problems with their research design, however, and found that they could not assess learning outcomes because "there are no standardized tests of chemical engineering knowledge" (Felder et al., 1997, p. 1287). Also, the experimental group and the comparison group were separated in time by two years and thus the researchers were concerned about a possible history effect (Felder et al., 1997). These two problems prevented the researchers drawing any meaningful conclusions about the effects of innovative teaching techniques on student performance and retention of knowledge.

To avoid running into the same pitfalls in this research, one particular topic was chosen as the specific domain for this study. Truss analysis was chosen as that topic for three reasons: truss analysis is an easy topic to assess, truss analysis is a topic that is covered in many IT packages, and truss analysis is a tough topic for many students. Each of these points are discussed in greater detail below.

First, truss analysis is a relatively easy topic to assess. A truss is a system of long, straight members that are connected in triangular configurations and loaded at the joints so as to develop a structure composed entirely of two-force members. Standard Newtonian Physics and trigonometry are used to analyze trusses. Performance and retention can be easily assessed by administering a test that contains two or three trusses and asking the students to solve for the member forces in the truss. Qualitative questions about truss assumptions and forces could also be asked and assessed rather simply.

The second reason for choosing truss analysis is that this topic is included in most software programs developed for use in statics as well as in tool-type structural analysis programs. This is beneficial because it allowed the researchers to choose between a wide range of software titles and, as mentioned previously, having a number of software titles to choose from allowed the researchers to choose one that met the goals of the research without having to spend time in developing new software.

Finally, and perhaps most importantly, truss analysis was chosen because it is a topic in which many students struggle. Interviews with instructors prior to beginning the project revealed this to be the case. Up until the truss analysis topic, the course focuses on analysis of single members only and the transition from single member analysis to analyzing systems of members can be difficult. Furthermore, instructors suggested that students have a hard time with trusses because they fail to see how the forces are distributed across the entire truss and tend to focus on individual portions of the truss instead. It was important to pick a topic that students were struggling with not only for the purpose of revealing a way to potentially help students but also because if a topic was chosen that all students excelled at, the intervention would be of little value and impossible to assess.

It is not within the scope of this document to describe trusses and truss analysis in detail, nor is a complete understanding of trusses necessary for an appreciation of this project. There are many excellent sources on trusses (i.e. Meriam & Kraige, 2002, Hibbeler, 2001, and McGill & King, 1995); repeating the information here would not be pertinent.

In summary, statics was chosen as the course in which to apply the research intervention because it is a low-level class that students take early in the curriculum, it allows for a large sample population, there are a number of readily available software titles for use in the course, and the results of the experiment can be generalized to a large population. Truss analysis was chosen as a specific topic within statics on which to focus the intervention because it is an easy topic to assess, it is included in many instructional and tool-type programs, and it is a difficult topic for some students to grasp.

3.5 Instruments

As shown in Figure 3.1, there were four separate tests used to measure three dependent variables. These testing instruments were developed with the input of faculty members who teach statics and truss analysis. The instruments were tested for reliability and validity during a formative study. The instruments, their development, the formative study, and the results of that study are all discussed in greater detail in Chapter 5. Please refer to that chapter for more information on the tests used in the study.

3.6 Apparatuses and Equipment

Outside of typical tools used in classroom teaching, the only equipment used for the study were the software titles that served as the independent variables and the computers that the software ran on. A detailed description of software titles considered as well as the titles that were chosen for use in this experiment is presented in the next chapter; please refer to Chapter 4 for more details on the software used. The computers used were Pentium III© based PC's running Windows 2000© operating systems with headphones available for student use.

3.7 Research Design

3.7.1 Research in Naturalistic Settings

In an effort to ensure a high degree of external validity, educational research is often performed in a naturalistic setting: the classroom. Semb and Ellis (1994) recommend conducting classroom research to conducting laboratory research when studying retention because it represents a more real-world setting for investigating memory. They point out that classroom studies focus on the relationship between manipulated variables and retention of knowledge and are, as a result, more functionalist in nature, as opposed to laboratory studies which tend to focus on the cognitive nature of memory. This study was intended to be functional; to determine whether or not software use enhances performance and retention and thus suggest its continued or discontinued use in the engineering classroom. Therefore, this study was conducted in the classroom.

Tunnell (as cited in Whitley, 1996) identified three dimension of naturalism, behavior, setting, and treatment. This project was designed to achieve a great degree of naturalism by meeting, to an extent, each of these three dimensions. The behavior dimension refers to the dependent variable being studied; this study satisfied this dimension by using typical assessment measures (i.e. take home assignments and in-class quizzes) as opposed to artificial measures such as self-evaluations. The setting for the study was also very natural, students are regularly in the classroom for the purpose of learning and as such it was a very natural place to study learning and retention. The treatment dimension may not have been natural for this particular group of students in this particular class, but satisfies the dimension nevertheless. The instructors may not have planned on using software in this class prior to this study and may not have used it for any other topics during the course, but many engineering instructors in many classes do encourage and or require students to use software. Requiring students to use software as part of the treatment in this project, therefore, was not an overly artificial treatment condition.

While research in natural settings does have advantages, external validity being the most obvious and important, there are disadvantages as well. The greatest disadvantage to conducting research in a natural setting as opposed to conducting research in the lab is the lack of control in the natural setting. The most important control that is sacrificed is random assignment of subjects.

In an effort to maintain naturalism, this research project, like many others, used naturally occurring groups of subjects rather than randomly selected/assigned subject groups. This is not uncommon in educational research. Son and VanSickle state "random assignment of subjects to experimental and control groups is generally impractical in…school research, because it is very disruptive of the normal classroom procedure and organization," (2000, p. 98). Son and VanSickle (2000) decided to use quasi-experimental design, specifically the nonequivalent comparison group design, for their classroom research.

3.7.2 Quasi-experiments and Nonequivalent Comparison Group Design

A quasi-experimental design is one in which random selection or assignment of subjects is not accomplished. Thus, all research using predefined groups of subjects is, by definition, quasi-experimental. Whitley (1996) explains that the nonequivalent control (or comparison as stated by Son and VanSickle) group design is the most common quasi-experimental design and is useful when random assignment is impractical.

The nonequivalent control group design is one in which two or more groups of subjects are studied, where one group acts as the comparison group and one or more groups act as the experimental groups (Whitley, 1996). These groups receive different treatments and are measured and compared for differences. The groups are assumed to be unequal in this design because only through random assignment can non-experimental factors (such as those listed in Section 3.2.3) be assumed to be equal among the groups.

3.7.3 Design Weaknesses

Whitley (1996) lists two major problems with this designe, preexisting differences and selection bias.

3.7.3.1 Preexisting Differences

Preexisting differences were mentioned in Section 3.2.3 and would exist if groups differed in knowledge or ability prior to the intervention. Whitley (1996) recommends pretesting as a way of ruling out preexisting differences. Comparing pretest scores, as well as other demographic data such as grade point average, ensure that the groups are similar in areas that could effect measurement of the dependent variable. As such, a pretest was part of the experimental design for this research project as well as the collection of demographic data. These data and scores were compared across groups and the results are presented in Chapter 7. The design and evaluation of the pretest is discussed in Chapter 5.

3.7.3.2 Selection Bias

Selection bias occurs when subjects are assigned to groups non-randomly (e.g. voters assigned to certain districts based on geography) or self-select the group they belong to (e.g. students selecting which section of a course to take). For example, students may select a particular section because their friends are taking that section in which case they may have similar personal characteristics, which could potentially confound the experiment (Whitley, 1996).

Whitley recommends two ways of overcoming selection bias: replication and multiple naturally occurring groups randomly assigned to experimental conditions. Replication is simply conducting the experiment multiple times with different groups while the later option involves using multiple groups assigned, randomly, to the experimental and control conditions.

For this study, replication was chosen because it was more feasible to conduct the experiment using three sections of CEE 2020 in two consecutive semesters than it was to use nine sections in one semester. There are a few reasons for this. First, convincing two or three instructors to allow the experiment to be conducted in their classes was easier than convincing six or seven (some instructors teach multiple sections). Second, if the instructor bias was to be overcome through the use of a common guest lecturer, as discussed in Section 3.2.3, it would have been nearly impossible to find such a person to teach nine sections of the course, even if it was for a limited amount of time.

3.7.4 Summary

In summary, the nonequivalent control group design was chosen with the intent of using entire sections of CEE 2020 as subject groups. Replication and pretesting were used to overcome some of the weaknesses of this research design.

3.8 Procedure

Implementation of the experiment began with the selection of subjects, continued through the intervention and concluded with the collection of the final data. This section will discuss the steps that were originally outlined for how the experiment was going to be conducted. Again, the actual implementation of the experiment differed slightly from the original plan; the actual implementation is detailed in Chapter 6 whereas the proposed procedure only is outlined, briefly, in this section. Explicit details of the originally proposed procedure are not given here so as not to confuse the reader when the details of what actually took place are given in a later chapter. The purpose in presenting this section is to reveal how the project evolved over time; throughout the remainder of this document changes to this procedure will be mentioned and so this section serves as the baseline for those changes.

3.8.1 Subject Selection

As mentioned in the previous paragraphs, subjects were to be selected as groups and not as individuals. The selection of groups was initially intended to be accomplished by matching groups containing similar population distributions of non-experimental variables such as gender, ethnicity, and age. Demographic information of this sort was available from the institute's registrar's office. Statistical tests (chi-squared) of the demographic data could reveal sections where no significant difference occurred between the distributions of the data and the three sections that were most similar would be chosen as the three groups for the study.

Participation in the study would be voluntary but measures were in place to motivate students to take part in the experiment. First, students would be told that they would be tested on the information presented both in the lectures and special sessions that took place during the intervention. Second, students would be told that they would earn extra credit in the course for completing the assessment instruments.

3.8.2 Intervention

Once the research groups were selected, permission from the instructor to intervene would be sought. Instructors would be informed that this intervention would include two lectures and one special session. The two lectures would include the material that was normally taught during the truss portion of statics: general truss information, method of sections, method of joints, and shortcuts. These two lectures would be taught by a guest lecturer to all three groups to exclude a possible instructor bias.

The special session is where the independent variable would be introduced. This session was intended to be a recitation session that students would take outside of the normal class time. Students could sign up for one of two offerings of the session targeted to their section. Students in all three groups would work on the same truss problems during this session and new material would not be presented. The truss problems would emphasize traditional analysis techniques as well as require students to determine the effects of adjusting certain truss parameters. This ability to work with the truss as a whole and intuitively identify such effects was identified by instructors as crucial. More information on this topic is presented in Chapter 4.

Students in the comparison group would work out the problems by hand and the guest lecturer would be available to answer any specific questions that students may have during this session. Students in the tool-type group would work on the problems with the help of a commercially available structural analysis program. Having selected a highly usable piece of software (see next chapter for more information) the guest lecturer would give a brief introduction on the software and how to use it to analyze trusses and would be available to answer questions about trusses or the software while students worked out the problems. Students in the content-type group would work on the problems with the help of a commercially available piece of software designed for use in engineering courses. Again, the guest lecturer would introduce the software and instruct the students briefly on how to use it. The instructor would remain present throughout the rest of the session to answer questions about trusses or the software.

3.8.3 Data Collection and Analysis

Four different measurements were to be taken during the study: the pretest, posttest, retention posttest, and long-term retention test. It was initially proposed that the pretest would be given during the first lecture on trusses and completed in-class. The posttest would be administered at the end of the special session and would be completed prior to the students leaving the session. The retention posttest would be administered by the guest lecturer ten weeks later in the semester, approximately just prior to the week scheduled for final exams. Again, it was intended that this retention posttest would be completed in-class.

Administration of the long-term retention test would require more effort because at twenty-five weeks, students were no longer in CEE 2020. As a result, students would be tracked, sent the long-term retention posttest via their school-administered e-mail account, and asked to return the completed instrument. An alternative to this was to administer the long-term retention test in one or more of the follow-up courses to statics such as engineering materials or mechanics of materials in which case the details of the collection could not be planned in advance but would depend on how the instructor of the follow-up course would be willing to participate. While participant mortality (failure of participants to take part in the follow-up study) would obviously occur in both cases, there is no reason to assume that the mortality rates would be differential or that one section would have a higher mortality rate than another.

Standard statistical tests and procedures would be used to analyze the data. Specifically, it was proposed that an analysis of variance (ANOVA) would be used to compare mean scores on assessment instruments from the different sections to determine if there was a significant difference. Should a significant difference be found, post-hoc tests would follow to determine the nature of the difference. Other tests, such as chi-square for example, would be used to break down the data and compare results based on demographics to determine if there was a significant difference in performance and retention based on non-experimental factors.

3.9 IRB Review

Following the approval of the proposed methodology by the research committee, the research plan was then sent to the Institutional Review Board for their review. Any academic research project that uses humans as subjects must be reviewed by the Institutional Review Board (IRB) at the school where the research is being conducted. This is done to ensure that rights and welfare of the participants are protected and that the research is conducted according to basic ethical principles (Office of Research Compliance, 2001). Most educational research wherein the activities involved do no go beyond normal classroom procedures fall into the *exempt* category where a full review of the research protocol is not required.

Because the subjects of this research were students, and therefore humans, an application had to be filed with the IRB at Georgia Tech, but the application requested an exempt status because the research activities were not going to go beyond normal classroom procedures. Exempt status was granted, see figure 3.2, on the condition that the principal investigators had successfully completed Human Subjects Training. Both principals had previously been certified to work with human subjects as required by Georgia Tech and proof of this certification was forwarded to the IRB. Once exempt status was granted, no further review by the IRB was required unless significant changes were made to the research plan. Although some changes were made throughout the research project, see chapter 6, none of these changes were significant and no further review by the IRB was requested.

Georgia Institute of Technology

Office of Research Compliance
Atlanta, Georgia 30332-0420 U.S.A.
PHONE 404-894-6944
FAX 404-385-0864
irb@gatech.edu
iacuc@gatech.edu

Protocol H01147
Funding Agency: National Science Foundation
Review Type: Exempt
Title: "Assessment of Technology-Based Instruction Impacts Toward Supporting Knowledge Retention of Engineering Undergraduate Students"

December 5, 2001

Dr. Nelson Baker
Civil and Environmental Engineering
Mail Code 0355

Dear Dr. Baker:

The Institutional Review Board (IRB) has carefully considered the referenced protocol. The proposed procedures are exempt from further review by the Georgia Tech Institutbnal Review Board.

Thank you for allowing us the opportunity to review your plans. If any complaints or other evidence of risk should occur, or if there is a significant change in the plans, the IRBmust be notified.

If you have any questions concerning this approval or regulations governing human subject activities, please feel free to contact Dr. Phillip Sparling, IRB Chair, at 404/8943402, or me at 404 / 894-6949.

Sincerely,

Barbara S. Henry

Barbara S. Henry
IRB Administrator

Enclosure
cc: (without enclosure)
Dr. Phillip Sparling, IRB Chair
Janis Goddard, OSP

Figure 3.2 Letter Granting Exempt Status from the IRB at Georgia Tech

3.10 Summary

This research project was designed to determine if learning and retention could be improved through the use of technology and if there was a difference in learning and retention when using tool-type technology versus content-type technology. To determine this, a quasi-experimental design was chosen in order to preserve naturalism in the study. In this design, entire sections of CEE 2020 would be chosen as subject groups and would undergo one of three treatment conditions during the intervention: no software use, use of a tool-type software, or use of content-type software. Learning would be measured with a posttest immediately following the intervention, retention would be tested with the same test at ten weeks, and long-term retention would be measured with the same test at twenty-five weeks. Results from these tests would be analyzed using accepted statistical methods.

An important step in the methodology that was mentioned but not detailed in this chapter is the process that was used to select the software that students would use during the intervention. The selection of both a tool-type and a content-type software was a lengthy process and was essentially a project within a project. As such, it is detailed in the following chapter.

CHAPTER 4

SOFTWARE SELECTION

The main focus of this research was to determine the impacts of educational technology on short and long-term retention. As such, a natural part of the research was to choose which software was going to be used in the classroom as part of the experiment. Developing new software for students to use was beyond the scope of this research especially when a number of different titles pertaining to trusses were already available for use in the classroom. Deciding on or selecting a standard by which to evaluate different software titles and eventually choosing one was a necessary task.

4.1 Predictive Evaluation

Heller states, "instructional software, like all other educational material, should be evaluated before it is used in the classroom or research project," (1991, p. 285). Squires and Preece (1999) agree and refer to this type of evaluation as predictive evaluation. Predictive evaluation has been a popular topic since the arrival of personal computers and educational software in the 1980's. Many people in all fields of education agree upon the need for predictive evaluation; the method for evaluating, however, is much debated.

4.1.1 Background

A number of different methods for performing predictive evaluation of educational software have been proposed over the last two decades. Perhaps the most popular method is the use of checklists. Checklists are often employed because they can be objective, easy to use, and general enough to be used for a wide range of software contexts and titles. Historically, software evaluation checklists have been developed by numerous individuals and groups including MicroSIFT (Microcomputer Software and Information for Teachers), EPIE (Educational Products Information Exchange) and NCET (The National Council of Teachers of Mathematics) (Heller, 1991). Checklists usually attempt to encourage evaluators to think about technical, usability, and interface issues as well as educational and content issues. The MicroSIFT and NCET checklists are reproduced in Figures 4.1 and 4.2 respectively; these are two typical examples of checklists that have been widely used in the past.

1. **Which computer system will the disc run on?**
2. **Will your computer system do justice to the illustrations?**
3. **Is the operation by keyboard or mouse, or both?**
4. **Can we have the disc for a trial period?**
5. **Is the language and spelling on the disc Queen's English or American English?**
6. **How much bias is there in the content of the disc?**
7. **Is printing out easy and intuitive?**
8. **Can the selected material readily be down-loaded to disc?**
9. **Can subsections of the disc be searched?**
10. **Is the software to control the CD-ROM on the disc itself or is it supplied on a separate floppy disc?**
11. **Does the software manage memory resources well?**
12. **What search procedures are available?**
13. **What is the language level on the disc?**
14. **Is the user interface tolerant of typing and spelling errors?**
15. **Can you select exactly what you want to print out or save to disc?**
16. **Are there any supporting features?**
17. **Can the illustrations be printed out?**
18. **Can images be readily transferred?**
19. **Is there a sound capability to accompany the pictures?**

Figure 4.1 NCET CD-ROM Checklist (Squires & Preece, 1996)

Another type of predictive evaluation is the Perspectives Interaction Paradigm (McDougall & Squires, 1995a). This method requires evaluators to consider the interactions that will take place between the teacher, student and designer while the software is being used. The teacher-student interaction considers the roles that both will play during the software use. The teacher-designer interaction suggests that the software content should correlate with the course curricula. The student-designer interaction is meant to investigate the learning theory upon which the software has been designed. Aside from these three types of interactions, this method has no specific questions or areas of investigation. It is a very open-ended type of evaluation that requires much thought and effort on the part of the evaluator. Despite being hard to learn and apply, however, users have found it to be somewhat beneficial (McDougall & Squires, 1995b).

The Jigsaw Model is another method proposed for evaluating software (Squires & Preece, 1996). This method attempts to incorporate usability and learning issues into a single model; it breaks the use of the software into four areas: 1) specific task concepts, 2) general task concepts, 3) system features, and 4) application features. The model suggests that evaluators should examine the relationships between areas 1 and 2 to evaluate learning and areas 3 and 4 to evaluate usability. Evaluators are then to look further at the relationship between all areas in order to assure that both usability and learning are not only assessed but also examined as one whole integrated task. As with the Perspectives Interaction Paradigm, this is an open-ended and subjective approach with little structure.

CONTENT

1. **The content is accurate**
2. **The content has educational value**
3. **The content is free of race, ethnic, sex, and other stereotypes**

INSTRUCTIONAL QUALITY

4. **The purpose of the package is well defined**
5. **The package achieves its defined purpose**
6. **Presentation of content is clear and logical**
7. **The level of difficulty is appropriate to the target audience**
8. **Graphics/colour/sound are used for appropriate instructional reasons**
9. **Use of the package is motivational**
10. **The package effectively stimulates student creativity**
11. **Feedback on student responses is effectively employed**
12. **The learner controls the rate and sequence on presentation and review**
13. **Instruction is integrated with previous student experience**
14. **Learning is generalisable to an appropriate range of situations**

TECHNICAL QUALITY

15. **The user support materials are comprehensive**
16. **The user support materials are effective**
17. **Information displays are effective**
18. **Intended user can easily and independently operate the program**
19. **Teacher can easily employ the package**
20. **The program appropriately uses relevant computer capabilities**
21. **The program is reliable in normal use**

Figure 4.2 MicroSIFT Checklist (Squires & Preece, 1996)

Reeves and Harmon (1994) recommend the use of a method that evaluates software on 24 dimensions: 14 pedagogical dimensions and 10 user interface dimensions. The pedagogical dimensions include areas such as epistemology, underlying psychology, and role of instructor. The user interface dimensions include areas such as ease of use, navigation, and screen design. In this method the reviewer is to examine where the software falls between two extremes on each of the 24 dimensions. An example of a completed review of the pedagogical dimensions of a particular piece of software is given in Figure 4.3. This evaluation method is qualitative, as opposed to the more quantitative checklists, and is meant to assure the reviewer that the software being evaluated follows his/her educational perspectives and is useable.

There are numerous other methods also. An heuristic approach to software evaluation is recommended by some as a way of integrating usability and learning (Squires & Preece, 1996). Jones et al. (1999) developed a Context, Interactions, Attitudes and Outcomes (CAIO!) method that was designed to account for the context in which the software will be used. Still, others designed methods for specific contexts such as health education (Premkumar, Hunter, Davison, & Jennett, 1998).

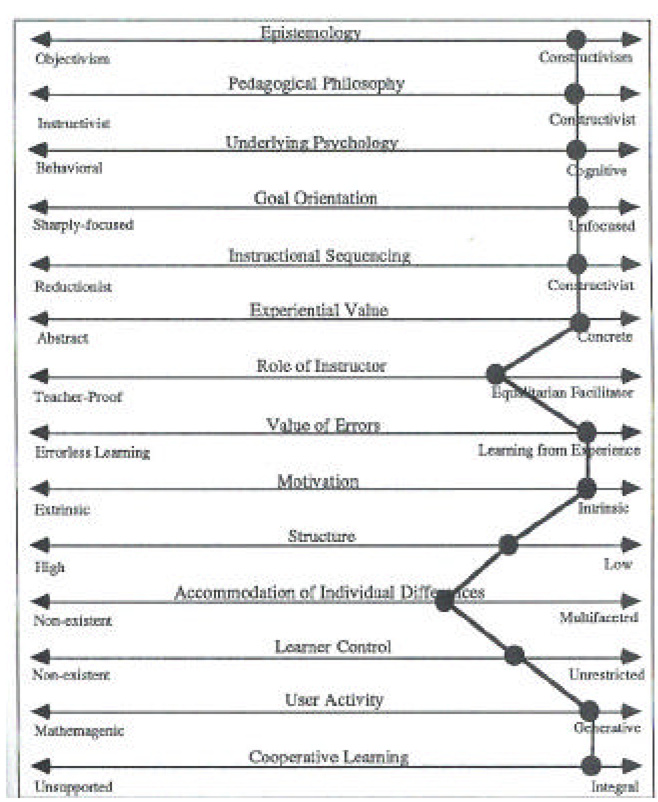


Figure 4.3 Completed Evaluation of Pedagogical Dimensions (Reeves & Harmon, 1994)

The methods mentioned above are just a few examples of the many types of evaluation techniques that have been published for use; a complete and exhaustive list is outside the scope of this work. There are a number of reasons that so many different methods exist as opposed to one single standard method. One reason is that evaluators debate over how in depth an evaluation tool should be as opposed to how easy the tool is to use. Checklists, for example, are very easy to use and can be completed by almost anyone. This type of evaluation is desirable for people such as pre-college teachers who are not given time outside the classroom to evaluate or review educational software titles (Heller, 1991). On the other hand, the Perspectives Interaction Paradigm may lead to more in-depth reviews but actually requires instructors to attend a two-day training seminar to learn how to use the method properly.

Contextual difference is another reason that various methods are used. For example, the MicroSIFT checklist, which is supposed to be very generalizable, was praised as being the first method to mention anything about race and gender stereotypes (Heller, 1991). Not all contexts, however, are sensitive to stereotypes. Whereas a piece of software that traces the history of slavery may need to be checked for biases, engineering contexts usually talk about things rather than people. Indeed, none of the software titles considered in this project mentioned people at all, much less stereotyped against a particular group of people. Thus, context must be considered when choosing an evaluation method.

Another similar reason is that instructors may want to ask specific questions or investigate certain aspects of the software that are not part of the evaluation method. This is why many instructors develop evaluation methods for their own needs and situations.

This is main reason that Premkumar et al. developed their evaluation tool; because none of the readily available software evaluation "tools address[ed] the unique needs of health education" (1998, p. 244).

The presence of numerous evaluation methods, however, is not necessarily undesirable. In none of the papers mentioned here did the authors suggest a method that could be generalized to all types of educational technology. Heller mentions numerous processes and forms with their relative strengths and weaknesses but refrains from recommending some at the expense of others (1991). The method used to evaluate software should relate to the intended purpose, context, and setting wherein the software will be used. With this idea in a mind, an evaluation method was designed specifically for this research project.

4.1.2 Development and Testing of Evaluation Form

Initially, an evaluation form was developed for this project but it was not used due to poor results on reliability tests. This section describes the form and the problems associated with it.

This evaluation form was built around Olcott's (cited in Palloff & Pratt, 2001) five "I's" of distance learning: *interaction, introspection, innovation, integration, and information*. Olcott is cited as suggesting that no single technology can fit all learning situations, and that the technology that is chosen for a particular use should address each of these five "I's". This framework was modified slightly for use in this project; *interaction* in distance learning refers to the communication between instructors and students and the technology that facilitates that. In educational technology, interaction occurs between the user and the program and it usually occurs via the program's interface. Thus *interaction* was replaced by *interface* to address not only interaction issues but usability issues as well. Five statements were written for each category and evaluators were to rate each statement as it applies to the software using a Likert type scale from one to five where one represents agreement and five represents disagreement. Some of the statements were based on heuristics suggested by Squires and Preece (1996) while others were developed specifically for the context of this project. The complete form can be found in Appendix A; some example statements from the form include:

> The software interface is easy to use and requires little cognitive demand.
>
> The user is allowed to control the pace of the interaction.

The form was tested on two truss sections of Gramoll's online e-book (2002) by two different raters one of whom was a Statics instructor while the other was a graduate student who had taken Statics a number of years ago. The form was intended and developed to be objective; ideally the form could have been filled out by anybody and yield similar results. When tests of reliability were run for the two raters, the result was an inter-rater reliability of nearly zero. In fact, the 95% confidence interval included both positive and negative reliability coefficients. Using more raters could have possible improved the results and perhaps identified one of the two raters as an outlier, but comments from the raters seemed to imply that the form may have been rigid and possibly too objective. The form did not account for some of the differences in application as well as contextual differences between the titles of software being considered. For these reasons the form was not used and a new approach to deciding which software titles to use for this project was employed.

4.1.3 Goal Centered Approach to Evaluation

Comer and Geissler (1998) explain that rather than adopting an existing checklist, evaluators should first and foremost explicitly define the goals of the software. Following this advice, three Statics instructors met with the school's Associate Chair for Information Technology and discussed using software in Statics and what they would like the software to be able to do. A number of goals came out at that meeting: it was decided that the software should be self paced, easy to use and navigate through, and visual in nature (because trusses themselves are visual, it is nearly impossible to solve a truss without a diagram). It was also decided that the software should correctly integrate with the curriculum or that it should use subject matter, vocabulary, and solution methods that are common to the study and practice of civil engineering. It was unanimous, however, that the most desirable attribute of a piece of software in this context would be the ability for the user to adjust certain parameters of a truss, such as the loading or the supports, and see the results visually. This would allow the user to determine, for example, how changing the direction of a lateral load changes the internal forces of the truss members. It was believed that experimenting in such a manner would help the students to see how the members interact with each other, to see the truss as a whole rather than just as a number of calculations that need to be trudged through. This was the main feature that the instructors wanted to see from a software package and, as it turns out, was also the deciding factor.

4.1.4 Suggestions from Learning Theory

The software evaluation approach used in this project was further developed based on suggestions from learning theory. This section will briefly discuss modern learning theories and how they contributed to the software evaluation process.

There are a number of different theories about how learning is actually accomplished. New developments in learning started over half a century ago with rise of behaviorism (Sulbaran, 2002), which was built upon and followed up by instructional design theory (Dick, 1992). In both of these theories, the instructor possesses knowledge and skills, which he or she then passes on to the students. In opposition to these theories is the theory of constructivism, which states that knowledge cannot be passed on or even taught but must be constructed by the learner. Many leaning models have evolved from, or in opposition to, constructivism. Some of these include constructionism (Papert, 1991), social constructionism (Petraglia, 1998), situated cognition (Brown, Collins, and Duguid, 1989), cognitive flexibility (Spiro, Feltovich, Jacobson, and Coulson, 1992), and case-based reasoning (Kolodner, 1997). These theories differ in the recommended approach to teaching and learning in major and minor ways depending on the theory to which one subscribes. There are some, however, who believe that differing theories of learning can be integrated. Greeno (as cited in Sulbaran, 2002) proposes options for integrating behaviorism, constructivism, and situated cognition. Dick (1992), on the other hand, acknowledges major differences between constructivists and instructional designers but does mention that both sides could be improved by learning more about the other.

The purpose of this section is not to describe or promote any or all of these learning theories but to explain how tenets of these theories were used in evaluating instructional software. Despite the many different theories of learning, there are several common practices among most of the more popular models. Most agree that authentic contexts, activities, or problems are required to increase learning, retention, and transfer. The amount of real-world context necessary for good instruction varies between theories but most agree that it is necessary. Many theories also agree that students should be allowed to experiment within these real-world contexts. Open-ended models, problems, or environments that can be manipulated, created, or experimented on allow the students to construct their own knowledge about a particular context. The authenticity of this context facilitates in the transfer of this knowledge to future situations.

Based on these learning theories, it was postulated that for a piece of educational technology to be pedagogically sound it must include the aforementioned characteristics. Namely, it should include real-world scenarios, situations, problems, and environments. Also, students should be able to experiment within these real-world contexts to construct their own knowledge. Furthermore, Mayer and Chandler (1993) cite evidence and experiments that show that multimedia explanations, consisting of narration and animation, results in improved performance and problem solving transfer. Narrated animations have been understood to be effective because they reduce cognitive load by utilizing the student's auditory channel as well as the visual channel thus leaving more room in the cognitive working memory to process the information. This use of auditory information in conjunction with visual presentations has long been supported by leaders in engineering education as an effective use of technology (Baker et al., 1999).

4.1.5 Project Specific Goals of Instructional Software

The bulleted items below make up the complete list of goals that were used for evaluating and selecting software for this project. The list combines goals outlined by the instructors involved in the project with goals that are inferred from accepted learning theories. The list of goals, or desirable software characteristics, has been broken into a few categories for simplification.

- General Characteristics
  - Interactive
  - Self paced
  - Easy to use
  - Available and accessible
  - Multimedia explanations
  - Audio explanation
- Interface Characteristics
  - Easy navigation
  - Clear instruction
  - Clear links
  - Attractive interface
  - Standard or intuitive buttons, menus, and icons
  - Appropriate and helpful feedback
- Content/Context Characteristics
  - Integrates with the curriculum and profession
  - Appropriate vocabulary and subject matter
  - Real world problems
- Style Characteristics
  - Multimedia which adds to the presentation
  - Visual
- Constructivist Characteristics
  - Open-ended models or problems that can be designed, played with, or experimented on

There is some overlap in these goals or characteristics, for instance if a program does not contain clear instructions then it likely will not be easy to use. All of the instructors that were interviewed and surveyed agreed that, while all these characteristics were important, the single most important characteristic was the ability to manipulate objects and see how these manipulations affected the results. Specifically, with respect to trusses, the instructors wanted a program that enabled students to design trusses and to adjust certain parameters and see the results of these adjustments. This important characteristic is consistent with the learning theories outlined previously.

This list of goals and characteristics was then used to evaluate a number of different software titles that were developed for use in assisting with the teaching of truss analysis.

4.2 Content-type Software

The objectives of this research were to determine if performance and retention were significantly affected when using either tool-type or content-type software as opposed to regular classroom techniques. Thus, one tool-type software and one content-type software were chosen for use in the classroom as part of this project.

4.2.1 Software Titles

This section will describe each of the software titles considered as well as their individual strengths and weaknesses. While an attempt was made to evaluate all readily available software titles for use in statics, some lesser-known programs or programs that are not widely distributed may have been overlooked.

4.2.1.1 Multimedia Engineering Statics (MES)

MES is an ebook complement to an entire online courseware management tool for statics and dynamics build by Kurt Gramoll (2002) at The University of Oklahoma. The ebook is broken up into sections and subsections. Each subsection follows the same general outline and is made up of four web pages, the first of which introduces a real world case, problem or situation and is called the case introduction. The second web page is where theory is presented, the theory that is used to solve the case. The third page is the case solution; here the theory is put into action to solve the case that was previously introduced. These three pages all have multimedia content such as pictures, short movie clips, and audio explanations. The last page in each subsection is a simulation page. In the simulations, the students are allowed to do some constructivist activities by constructing or manipulating situations. For example, in one of the truss simulation pages, the students are allowed to build a small truss of their own design and load it up and see the results. They can then change the loads, supports, or members and see how this affects the results.

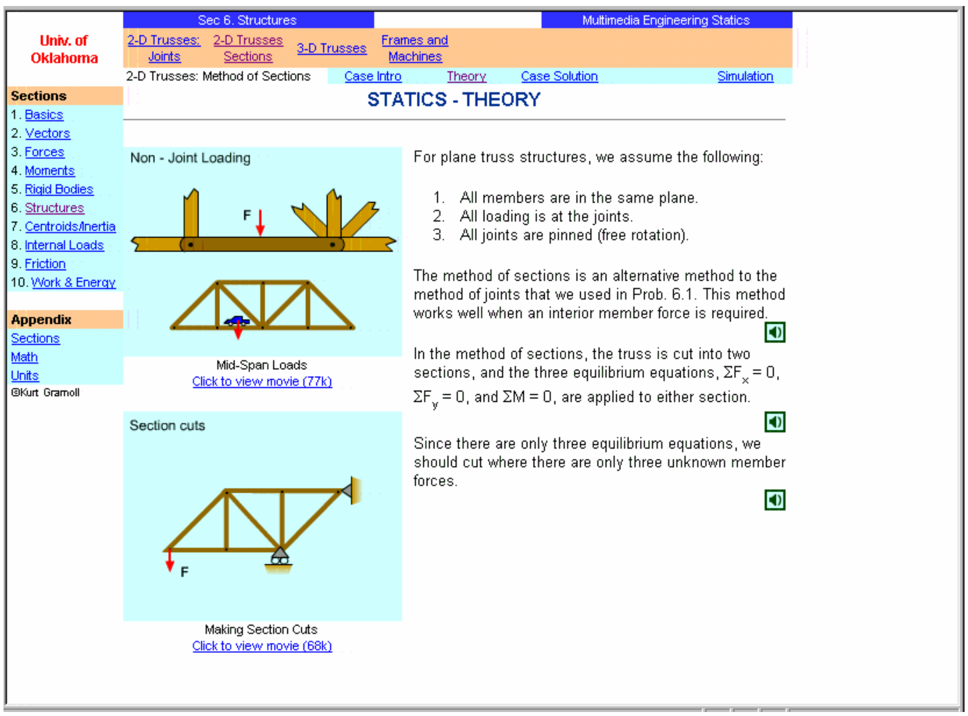


Figure 4.4 Screenshot from Multimedia Engineering Statics (Gramoll, 2002)

Figure 4.4 is a screenshot from one of the theory pages of this program; notice the section menu on the left and the four pages (case intro, theory, case solution, and simulation) menu below the subsection menu at the top of the page. This page uses two movies, three audio clips, and text to explain the theory behind the method of sections for truss analysis.

Many of the goals and characteristics outlined above were met or fulfilled through the use of MES. MES is interactive and self-paced. Being on the World Wide Web makes it available and accessible to everyone at no-cost; it is even available to users who are not subscribers of the entire courseware management system (which is also free but requires an instructor to set up a class). MES contains multimedia explanations and audio explanations throughout. The simulation pages on MES allow for constructivist-like exploration and experimentation. Textual explanations of theory and solutions are accompanied by helpful figures and equations. MES also includes a real-world problem in each of its subsections and uses terminology and methods that integrate well with the curriculum and the profession. The interface is attractive and the structure of the ebook is consistent.

The weaknesses of MES all relate to usability issues. The instructions for the simulation pages are not easily understood and some of the simulations are not very intuitive. Similarly, what little feedback is present for the simulation pages usually is not very helpful or constructive. While the structure of the ebook is consistent, it is not immediately intuitive and takes a few moments to determine the layout. This tends to make the ebook initially confusing but after some practice mastery is easily achieved.

4.2.1.2 MDSolids

MDSolids (Philpot, 2002) is the Premier Award (ASEE annual award for excellence in engineering education courseware) winning software developed by Timothy Philpot of the University of Missouri-Rolla. The software was intended for use in mechanics of materials courses though a few of the modules are useful for statics topics also. MDSolids is divided up into ten different modules that each solves a particular type of problem. The determinate beam module, for example, allows the user to design a simple determinate beam and the program draws the shear and moment diagrams. The user chooses the types and locations of supports and loads. Once the diagrams are drawn, the parameters can be manipulated and the diagrams updated. This allows the students to experiment with different conditions and see, visually, how changing parameters affects the results. The modules that are appropriate for use in statics are trusses and determinate beams.

MDSolids is a comprehensive and versatile tool for mechanics of materials courses. The program is equipped to solve nearly every type of problem that a student is likely to encounter during a mechanics of materials course and uses appropriate vocabulary and methods in doing so. MDSolids is available to students at minimal cost, $25, and can be downloaded from the company's website (Philpot, 2002). It is, however, only available for use on machines with Windows® operating systems. The program is self-paced and interactive but is mainly geared toward problem-solving and presents little actual content or theory and as such requires the students to have some knowledge prior to using the software. Most problems and solutions are accompanied by graphs and figures though there is no animated or audio content at all. Many of the modules are constructivist in nature and contain models that students can manipulate and study the results. Figure 4.5 is a screenshot from the truss module which was considered for use in this project. Note the instructions on the left, the truss (which is built by the user) in the center, and the results in the drop down menus at the top. Graphical results are also available.

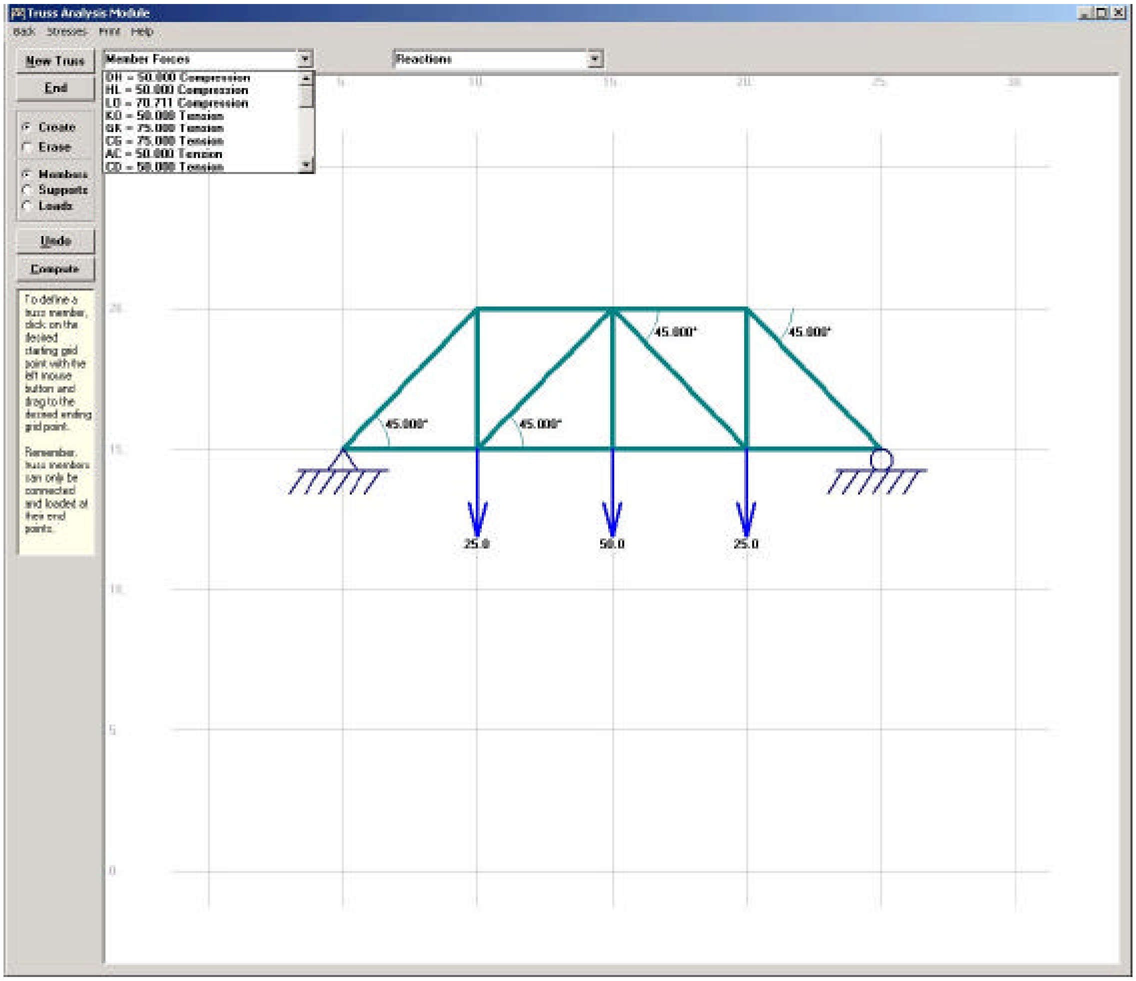


Figure 4.5 Screenshot from Truss Module of MDSolids (Philpot, 2002)

The comprehensive nature of MDSolids proves to also be a hindrance to the usability of the program. Each of the modules appears to have been developed as disparate, stand-alone products which, when brought together in this format, lead to inconsistencies across modules. An example of this is that some of the modules have help menus while others do not. Similarly, some start out with some information already given to students and input is simply required to complete the problem while other modules start out with blank screens leaving the user to explore the menus to determine what is to be done. Feedback varies from module to module and in some cases is very specific while in others it is very general and may or may not be useful to the student depending on his or her understanding of the material.

4.2.1.3 BEST Statics

This is a work in progress being developed by faculty and students in the Instructional Software Design Center at the University of Missouri Rolla. BEST (ISDC, 2002) is done in Flash, is available online, and has a continually growing database of theory, examples and problems. There are three separate sites, one each for statics, dynamics, and mechanics of materials. Each individual site then has a table of chapters and sections with links to theory, examples, and problems for each section within a chapter. The Flash animations make some complex visuals easy to understand and especially illustrate 3-D graphics well.

BEST is interactive and easy to use. Because it is online, it is available and accessible to everyone and is free to use. All of the theory, examples, and problems are animated and use multimedia explanations to help visualize the problems and situations. The interface is attractive and navigation between problems and sections is accomplished through the use of clear links. Some of the problems are interactive in that students are asked a question that must be answered correctly before moving on; students answer the questions by choosing one of the options presented to them (i.e. multiple choice type questions). The feedback given on these types of questions is specific and useful but is limited only the answer choice options available to the students. The content is authentic and uses language and methods that are consistent with the profession. Figure 4.5 shows a scene from one the example pages. Note the menus on the left and right; they are intuitive and easily understood. At this point in the example, the user is expected to answer the question correctly before moving on.

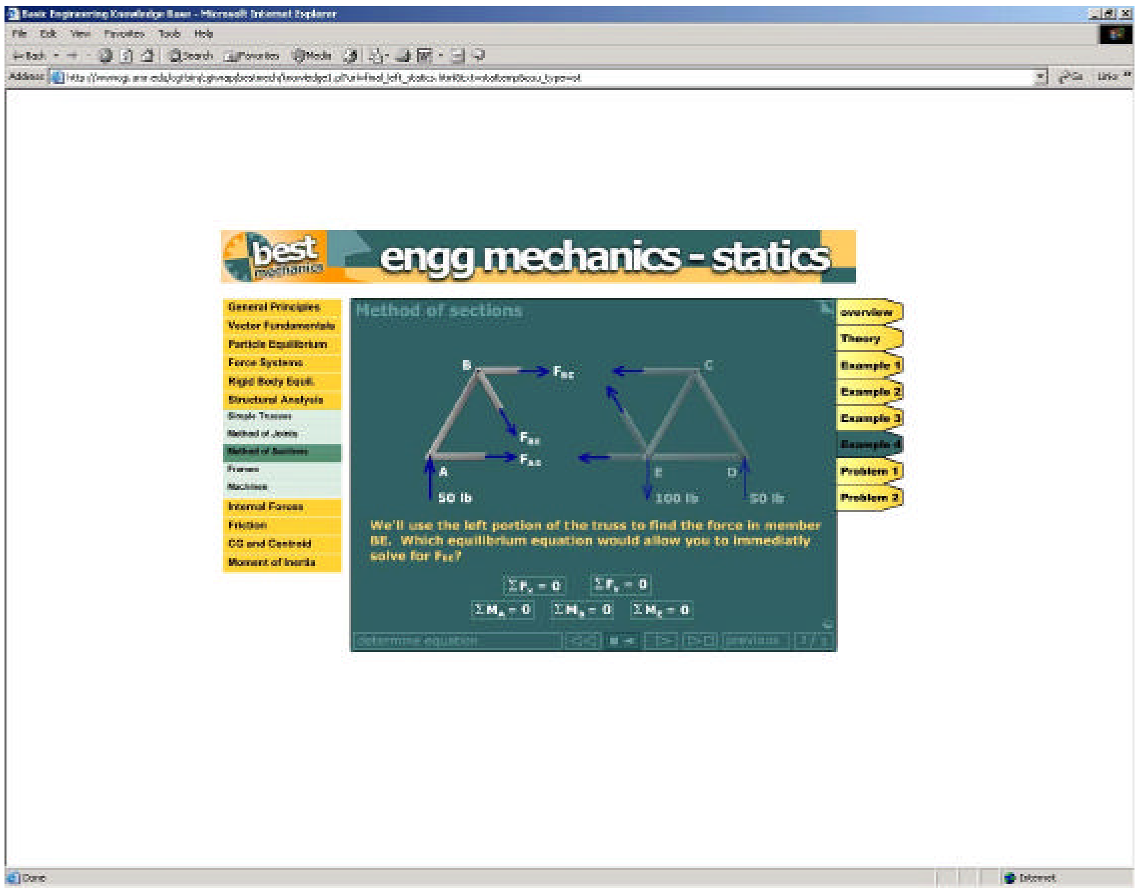


Figure 4.6 Screenshot from BEST Statics (ISDC, 2002)

In the context of this project, the major weakness of BEST is that it has no constructivist characteristics, which was identified by the faculty involved in the project as being the most important characteristic. The control panel, shown at the bottom in figure 4.6, is not immediately understandable but after a few practice runs becomes more usable. Unfortunately, the control panel does not come with any instructions. Another usability issue is that in some scenes the user clicks a button or an answer choice to move forward while in others the control panel is used to advance the presentation. It can be confusing sometimes as to which method to use. Simple cues at the end of each scene could make this process easier.

4.2.1.4 Multimedia Learning Environment for Statics (MLE)

This program was developed by Siegfried Holzer and Raul Andruet (2000) as an instructional aid for students in statics classes. The program presents theory, problems, and examples for a number of different statics topics. Navigation throughout the program is generally accomplished via links connecting one page of information to another. There are photos, diagrams, and simple animations throughout the program. There are also simple problems throughout that allow students to enter answers and then compare their results to the correct answers.

MLE is interactive, self-paced, and available for free on the World Wide Web, though it is only for use on Windows-based machines. Multimedia is accomplished through the use of static pictures that change or are updated with the click of a button; few animations are used. The program uses standard terminology and methods from the field of study and uses real-world problems in the examples and problems. Unlike some of the other programs, MLE uses many photographs to enhance the presentation and to anchor the topics in authentic settings. Figure 4.7 shows two screenshots from MLE, the first one showing a problem that is to be solved using method of joints and the second showing the analysis of one of the joints. Note the instructions at the bottom right of each screen; when Joint C from the first screen is clicked, per the instructions, the second screen pops up. MLE is an extremely informative and illustrative tool when used appropriately.

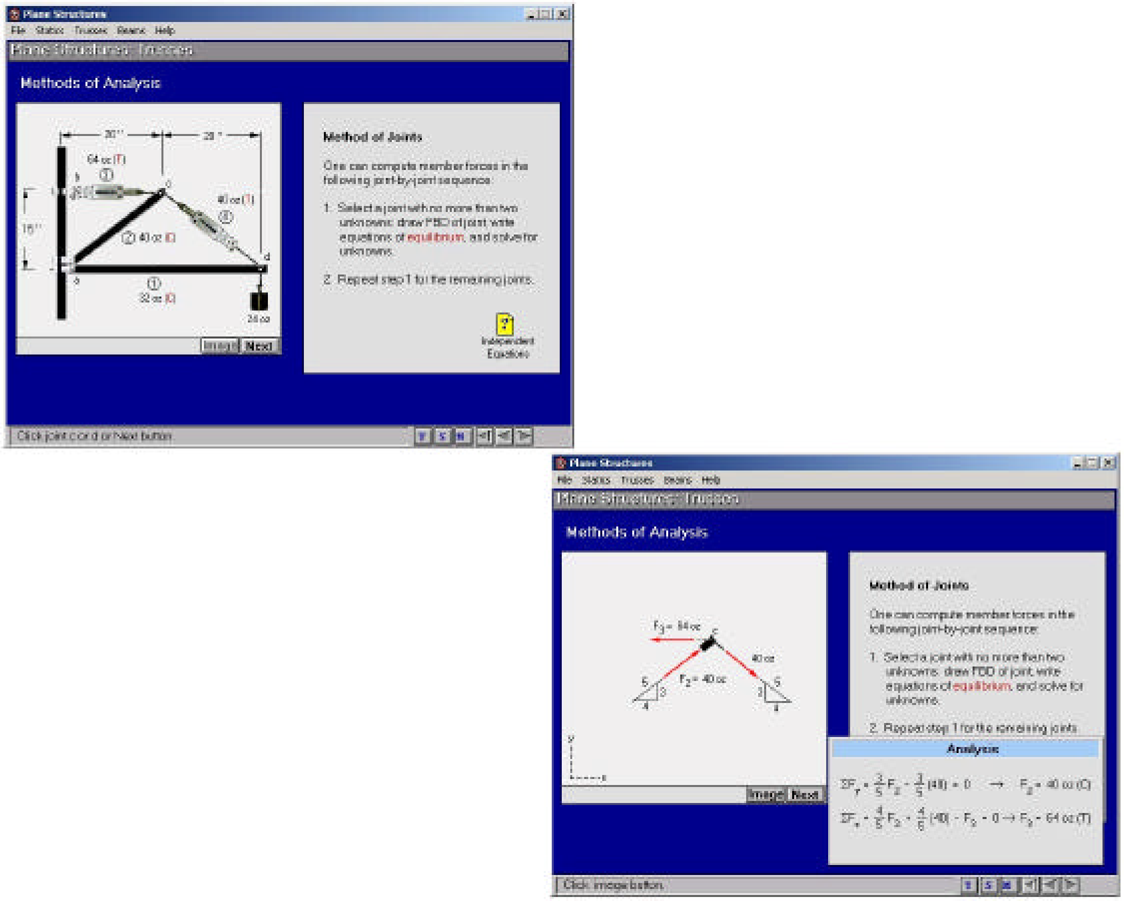


Figure 4.7 Screenshot from Multimedia Learning Environment for Statics (Holzer & Andruet, 2000)

MLE is promoted as being experiential as opposed to constructivist in nature. Although the two theories differ in some aspects, they both promote learning by doing. MLE attempts to achieve this by providing as little structure to the program as possible while still making it usable. Note that in the first screen in figure 4.6, there are eleven different buttons or links that the user can click to move from this screen to another. The small instructions at the bottom right provide some direction but still give three different options and sometimes the instructions are not as clear as this. The idea is that the student is to sit down with this software and experience it as opposed to merely progressing through it in a linear fashion. Students are to construct their own mental models of the information by following the links that they want to rather than being forced in a particular direction. If students use the software in this manner, exploring and experiencing the topics, it can be very beneficial. This form of experiential or constructivist learning, however, was found to be confusing by some instructors and students who evaluated the software. Users viewed the software as something to get through, similar to a homework assignment, as opposed to something they should experience. Users were unsure of where to go next and always had the feeling that perhaps they were missing some information by not clicking on the correct links or clicking in the wrong order. Furthermore, when students answered questions, the feedback provided was nonspecific (i.e. did not address students' answers directly) and simply in the form of correct answers.

4.2.1.5 Statics Tutor

Statics Tutor (DeVore, 2000) is a CD that contains explanations of statics topics and includes examples and problems for users to solve. For the most part, the explanations and examples follow a slide show format for presentation of information not unlike turning the pages of a textbook. The problems on the CD are slightly more interactive and involve a number of steps; the user must get one step of the problem right before moving on to the subsequent steps. This type of format allowed the program to give intermediate feedback on problems instead of simply telling the user whether the final answer was right or wrong.

The Statics Tutor is interactive, self-paced, and easy to use. It is published by Prentice-Hall and available at bookstores for around $40. The multimedia content consists mainly of simple sketches of common problems and some simple animated explanations. Because of the linear, textbook-like fashion of the CD, navigation is easy and links are clearly identified. Standard language and solution methods are employed and simplified real-world problems are presented.

The most important shortcoming of the Statics Tutor is that the solution to one of the truss examples was incorrect. The truss in question was very simple (five members, simply supported, one load, and symmetrical) and yet the answers were wrong. Allowing students to use a product which incorrectly demonstrates solution methods was clearly not acceptable. Aside from that, instructions on problems that students were supposed to solve were not immediately available. Seeing instructions either required clicking on a menu option or answering a question wrong three times in a row. Furthermore, the program did not allow for alternative methods of solution. For example, when finding an orthogonal component of a force, typing "cos30" may be counted as correct whereas typing "sin60" was considered incorrect even though they are numerically and conceptually the same thing. As with some of the other programs, feedback was problem specific and not answer specific but was provided at intermediate steps so that users can identify errors throughout the process as opposed merely finding out that they got the wrong final answer. DeVore's Statics Tutor contains no constructivist elements.

4.2.1.6 Statics Tutorial

The Statics Tutorial by Beer and Johnston (2000) is included as part of the New Media Version of their popular statics textbook. This tutorial is very similar to the Statics Tutor mentioned previously and uses slide-show format and simple animations to present theory and examples. This tutorial also has drill-and-practice quiz banks that students can take but, unlike DeVore's Statics Tutor, these problems do not include intermediate feedback but does provide students with correct final answers and solutions methods.

Statics Tutorial is interactive, self-paced, and easy to use. The CD comes with the current, New Media Version of the textbook, which costs about $125, but is not available as a stand-alone product. The program does use some simple multimedia animations to help with visualizations such as isolating a joint from a truss. The linear structure of the program is very intuitive with clear links and instructions. Feedback is in the form of correct answers and worked solutions. The problems, examples, and information integrate well with any statics course.

Statics Tutorial is good companion to the textbook but does very little that the textbook cannot. Aside from the simple animations and worked solutions, this product does not provide any additional value above and beyond the text. It does not allow students to construct or experiment with open-ended problems.

4.2.1.7 Working Model Simulations

Three popular textbooks include CDs with pre-set Working Model simulations. Bedford and Fowler's Statics Study Pack (2002) and Hibbeler's Statics Study Pack (2001) both include simulations using Working Model 2D and Beer and Johnston's New Media Version of their statics text (2000) includes a CD of simulations that use Working Model 3D. In all three cases, the simulations are based on examples from the book and allow the users to adjust the parameters of the model and see the results. Users are not, however, allowed to design their own models using these pieces of software.

Working Model is a very powerful visualization tool and is a popular computer aided engineering tool (MCS Software, 2003). The simulations that are provided with these textbooks are excellent and allow the users to visualize real world problems and scenarios. Users can adjust the parameters and see what actually happens to the system as a result in real time.

Unfortunately, the simulations are also very cumbersome. Viewing the simulations requires the installation of a Working Model Engine and a Working Model Viewer. Once this is done students can only open simulations that have already been developed and are not allowed to create new simulations. There are also only one or two simulations for each topic covered in the book so the amount of experimentation that students can do is very limited. The use of the working model viewer is not limited and few instructions are provided.

4.2.1.8 Problem Banks

Two textbooks, the aforementioned Bedford and Fowler Study Pack (2001) and the Hibbeler Study Pack (2001), both include a password to an online problems website. These websites include problems and solutions from former versions of the textbooks that students can try on their own to solve. The problems are categorized according to the chapters in the accompanying textbook. Both versions also include multiple choice and true/false questions to test theory and application; these are available for use without a password. The Bedford and Fowler companion site also has online homework capabilities where homework is completed and graded online and grades are reported to the instructors.

Figure 4.8 is an artificial example of a problem bank. The problem banks are just Web Pages with links to problems. If the link is clicked a new window opens up with the problem and another link which can be clicked for the solutions. Problem banks are easy to use and have clear and obvious links. Feedback is in the form of correctly worked solutions.

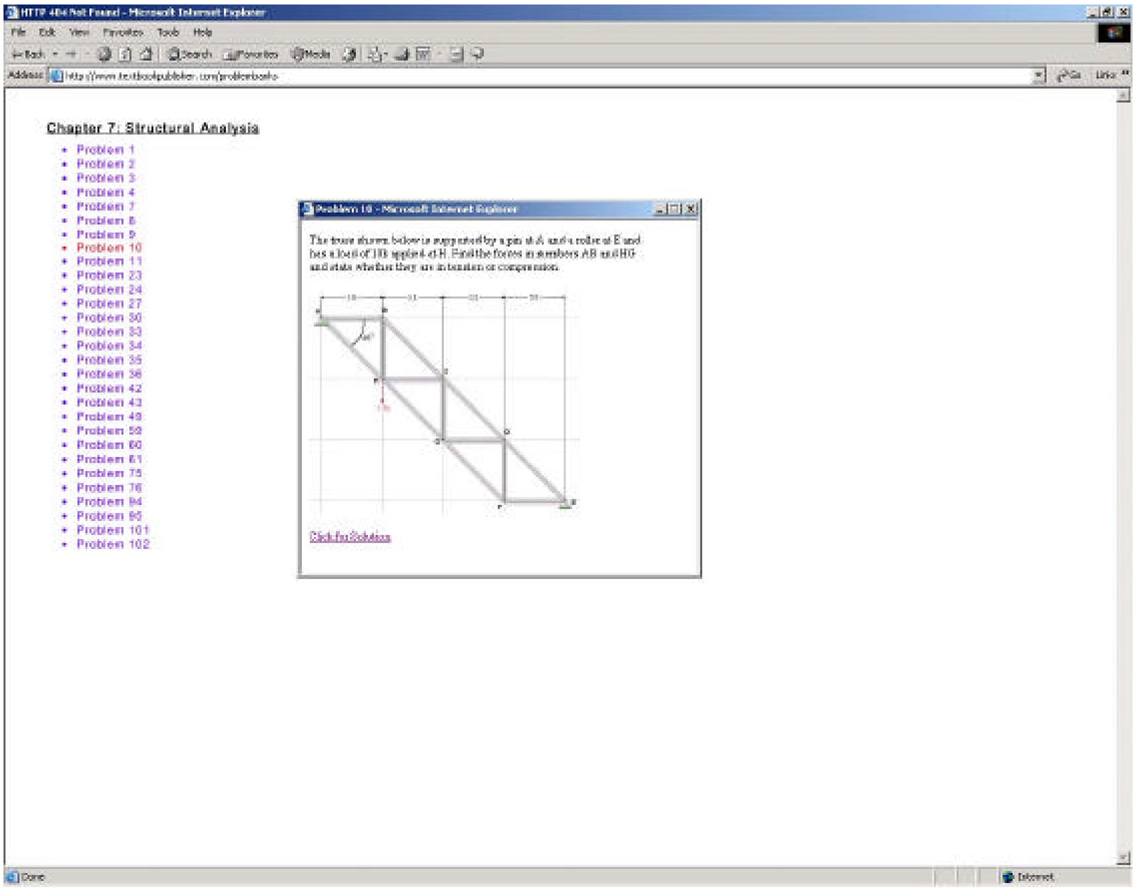


Figure 4.8 Example of a Problem Bank

Problem banks are not interactive and they do not have any multimedia explanations. Aside from the solutions, they provide little benefit beyond what a normal textbook has to offer and solutions may not be a benefit either as some instructors use different solution methods than the authors.

4.2.1.9 Shaping Structures: Statics

Although *statics* is in the title of this software, this title is more of an instructional program to teach structural analysis. This software gives excellent step-by-step examples of graphical solutions to eight different types of structural analysis problems. While trusses were one of the eight topics discussed, the method described in this software, the load line method, is one that is not normally taught in statics courses at the sophomore level but is reserved for more advanced structural analysis courses. This program is intended to accompany a text on graphical solution methods to structural analysis that is targeted to architects, not engineers.

Shaping Structures: Statics (Iano, 1998) is written in Adobe Acrobat and is self-paced and easy to use. The structure of the software is very linear and the user simply clicks a button to move forward and backward. The program does present theory and examples in a very graphical method despite the graphics being static. The instructions are clear and the problems are authentic.

The program contains no problems for the students to work out thought it is clear that users are to follow along with the examples on their own. The program also contains no constructivist-like activities. The main concern with this program, however, was the fact that it did not integrate well with the curriculum; it teaches a method that is not normally taught to sophomore civil engineering students.

4.2.2 Evaluation Results

Multimedia Engineering Statics (Gramoll 2002) was chosen as the content-type tool that was used in this project. Table 4.1 lists all the software titles that were evaluated, the evaluation characteristics, and how each title met the characteristics. Only two titles, MES and MDSolids, contained the type of constructivist characteristics that both statics instructors and learning theory deemed necessary. These two titles also met more of the evaluation characteristics than any of the other programs. While it was a tough decision to choose between these two exceptional programs, MES was chosen because it provided content as well as analysis. MDSolids, while it is designed to be an educational tool, is primarily used to help students in the analysis of mechanics problems whereas MES provides textual, audio, and multimedia explanations of background information, underlying theory, and solution methods in addition to analysis assistance. The analysis tools incorporated into MES were not as complex as those provided with MDSolids but they were sufficient for the needs of this project. MES would be even more desirable if the analysis tools could be more powerful without sacrificing ease of use. As discussed previously, MES is not perfect and did not meet all of the desired goals and characteristics but, for the particular needs of this project, it was the best tool evaluated. Also, the majority of the concerns associated with MES were interface-related and these concerns were easily overcome with just a few minutes of explanations.

Table 4.1 Software Titles and Evaluation Criteria

Evaluation Characteristics		Multimedia Engineering Statics	MDSolids	BEST Statics	Multimedia Learning Environment for Statics	Statics Tutor	Statics Tutorial	Working Model Simulations	Problem Banks	Shaping Structures: Statics
General Characteristics	Interactive	●	●	●	●			●		
	Self paced	●	●	●	●	●	●	●	●	●
	Easy to use	●	●	●	●	●	●		●	●
	Available and accessible	●	●	●	●					
	Multimedia explanations	●		●	●		●			
	Audio explanation	●								
Interface Characteristics	Easy navigation	●	●			●	●		●	●
	Clear instructions		●				●		●	●
	Clear links	●		●			●		●	●
	Attractive interface		●	●				●		●
	Standard or intuitive buttons, menus, and icons		●			●	●			●
	Appropriate and helpful feedback		●	●	●					
Content/Context Characteristics	Integrates with the curriculum and profession	●	●	●	●	●	●	●	●	
	Appropriate vocabulary and subject matter	●	●	●	●	●	●	●	●	
	Real world problems	●	●	●	●	●	●	●	●	●
Style Characteristics	Multimedia which adds to the presentation	●		●	●			●		●
	Visual	●	●	●	●	●	●	●		●
Constructivist Characteristics	Open-ended models or problems that can be designed, played with, or experimented on	●	●							

4.3 Tool-type Software

The evaluation of tool-type software titles followed a slightly different approach. Various structural analysis tools that are used by practicing structural engineers were considered for use as the tool-type software. Each of the software titles allowed the user to built, load and manipulate the parameters of two dimensional trusses and see the results of these actions. Thus, tool-types were not evaluated on the basis of functionality but on usability. Graphical input interfaces were a must because the experiment did not allow for time to teach the students a particular syntax for textual input. Unfortunately, cost was also a criterion for the evaluation of tool-type software titles. Some titles, even with an educational discount, would have cost thousands of dollars for a site license or the required number of individual licenses.

An attempt was made to consider some of the most popular structural analysis programs. There are, however, over sixty commercial structural analysis tools available for purchase (iCivilEngineer, 2003) and a comprehensive trial of each title was outside the scope of this project. Some titles were tested with a model truss while other evaluations were based on literature provided which included detailed information on input and usability.

One tool-type program was clearly superior to all other titles. As a result of this, a formal evaluation, similar to that done for content-type software, was not performed. A list of some of the programs considered for the project follows, but comprehensive information is not given here because a formal evaluation was not completed. Some of the programs considered were GT STRUDL (Georgia Tech-CASE Center, 2002), P-Frame (CSC, 2002), Dr. Frame (Dr. Software, 2002), ANSYS (SAS IP, 2002), VisualAnalysis (IES, 2001), STAAD Pro (Research Engineers International, 2002), RISA-2D (RISA Technologies, 2002), CADRE Lite (CADRE Analytic, 2002), ETABS (Computers and Structures, Inc., 2002a), SAP2000 (Computers and Structures, Inc., 2002b), and STRAP (ATIR Engineering Software, 2002).

The tool-type program chosen for use in this project was Dr. Frame. While Dr. Frame was not as powerful as some of the other structural analysis programs, it was significantly easier to use and sufficiently powerful for the needs of this project. To build a truss in Dr. Frame, students just point and click to insert members, supports, and loads. Once a stable truss was built and loaded, member forces were immediately reported to the user. Even more importantly, as the students changed parameters of or loads on the truss, the member forces were updated in real-time. For example, students could grab the leader of a force and drag the leader to increase or decrease the force and the member forces displayed would correspondingly increase or decrease at the same time. This functionality, along with the exceptional usability, was precisely what instructors were looking for in a structural analysis tool. The students could open the program and build and manipulate trusses within minutes. They could also experiment on those trusses and see the results of those manipulations immediately. No other structural analysis tool could accomplish this task with the ease that Dr. Frame did.

As a side note, though Dr. Frame was only used here to obtain member forces in trusses, with additional input (such as member properties) more detailed results (such as stresses) were available. Dr. Frame is not intended for determinate trusses alone either, but can be used for any type of two dimensional frame analyses under any type of loading. Dr. Frame does not, however, have finite element analysis capabilities that some of the large, more powerful programs utilize. Finite element methods of structural analysis are normally reserved for graduate courses and so Dr. Frame would be appropriate for any undergraduate courses involving structural analysis or mechanics.

4.4 Summary

Through a rigorous process involving learning theory and software evaluation literature review, interviews with instructors, and some comments by a few students, two software packages were chosen for use in this project. Multimedia Engineering Statics was chosen as the content-type software and Dr. Frame was chosen as the tool type software.

CHAPTER 5

FORMATIVE STUDY

Prior to conducting the actual experiment, a formative study was completed. This study, along with the precursory data that it produced, will be discussed in this section.

5.1 Purpose

In June of 2002, precursory data were collected from two sections of CEE 2020 taught in the summer; the same professor taught both sections. The primary purpose in gathering the precursory data was to assess the tests that had been developed; to assure that the questions on the tests were clearly stated, easily understood, and properly assessed the topics that they were designed to measure. From this data, a validation of the assessment instruments for this research was obtained. A number of different tests and questions were completed by the students in these two sections each of which will be described with results following the descriptions.

5.2 Instruments

Four different tests or problem sets were evaluated in this formative assessment: preliminary questions, pretest, posttest, and exam questions.

Preliminary questions tested information with which students should have been familiar prior to taking the truss portion of CEE 2020. Preliminary questions included questions about trigonometry, vector resolution, and equilibrium. These topics were identified by instructors of the course as requisite knowledge for analyzing and understanding trusses. In the context of the actual experiment, the results of the preliminary questions were intended to be used to compare the different sections prior to the experiment; to ensure that the sections each started the truss portion of the course with similar knowledge bases or skill levels.

Pretest questions were questions about trusses and were not expected to be answered correctly by most students because they had not yet been taught trusses. Within the context of the actual experiment, the pretest questions were intended to be used in conjunction with posttest questions, which were to be identical to the pretest questions, to obtain differential scores to determine how much students learned about trusses during the intervention.

The posttest, as just mentioned, was identical to the pretest and was intended to be used in conjunction with the pretest to get differential scores as a measure of learning.

Exam questions were questions that the instructor used to formally evaluate students' ability to analyze trusses. In the formative study, all students were given a problem on their midterm exam which required them to analyze a truss; see Figure 5.1.

5.3 Method

The purpose of the formative study was not to assess the students, but to assess the usability and reliability of instruments which would later be used to evaluate different groups of students. As such, no educational intervention took place during this initial study. Two sections of statics participated, but neither of them used any type of software in their study of trusses. Furthermore, since the same instructor taught both sections, it is assumed that the two sections received roughly the same lectures and materials.

Problem 5. Determine the forces in members *AB*, *BR* & *QR* of the truss below. Indicate whether the members are in tension or compression. Show all work.

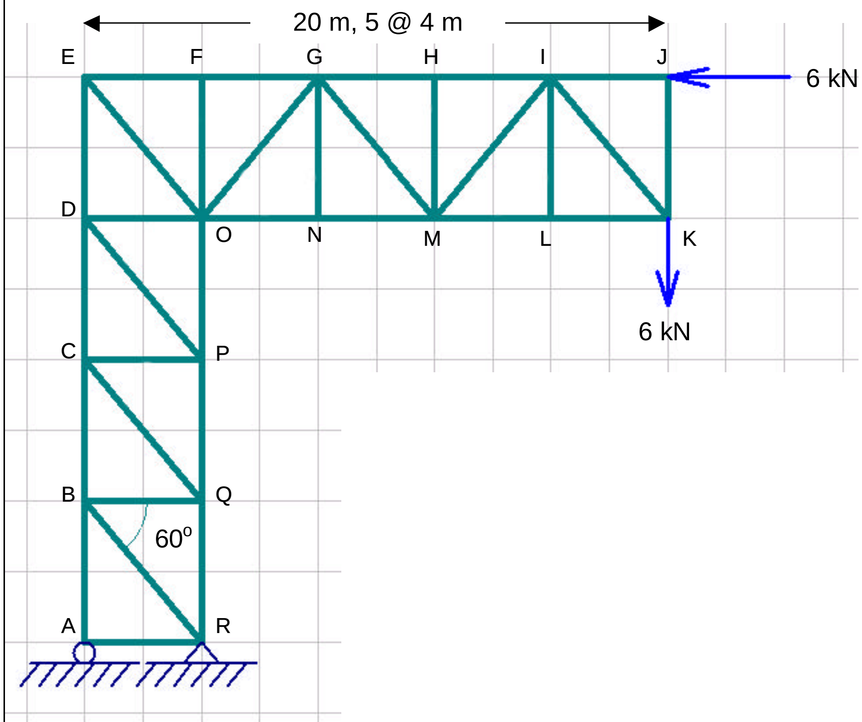


Figure 5.1 Truss Problem on Midterm Exam: Formative Study

The pretest and the preliminary question were to be given to the students before the truss portion of the course. These questions were given to the students as a take-home assignment. The students were informed that the test would not be graded but that they would receive bonus points for completing the test. They were further informed that some of the questions would be new to them and that they wouldn't be able to solve them and were thus encouraged to answer *I don't know* to any problem that, upon giving an initial try, they were unable to complete. Students were asked to not use books, notes, or any other resource to answer the questions.

It was decided in a meeting with the members of the faculty who teach CEE 2020 that shuffling the preliminary and pretest questions together to make one test would be best so that students would see a mix of familiar and unfamiliar questions. This was the approach taken in gathering the precursory data and the combined/shuffled test was given as the take home test. The instrument that included both preliminary and pretest data in shuffled form is in Appendix B.

The take home test was given on a Friday and students were asked to return it by following Monday in order to receive a 1% bonus added to their final grades. Unfortunately, due to conflicting time schedules and miscommunications, the students were given one truss lecture before taking the take home test but no homework had been assigned yet. Students were allowed to take as much time as they needed on the test but were asked to note the time they took and record it on the front page of the test.

After the instructor lectured on truss analysis, students completed the posttest. The intended methodology was to have the questions on the pretest be identical to the questions on the posttest. However, time constraints caused by a short summer semester only allowed for 20 minutes of class time to gather posttest data, which was not enough time to administer the full pretest again. Thus, one section of students was given half of the pretest/posttest questions and the other section was given the other half.

Students were also given a midterm once the truss lectures were completed. In addition to other equilibrium questions, the truss analysis question pictured in figure 5.1

was on the exam. The results of this exam question were analyzed for reliability in the event that this question might be used for assessments during the actual experiment.

The results of these assessments were analyzed using standard parametric and nonparametric statistical methods. The results, statistical methods, and analyses are discussed in the following section.

5.4 Results

The results of the various tests will be presented here followed by analyses of the results. Six out of 14 students from one section and 15 out of 29 from the other choose to complete the take home preliminary/pretest assignment. Again, they were allowed to take as much time as they needed but were asked to record the amount of time they spent on the assignment. A histogram of the amount of time students took to complete the test is given in Figure 5.2. Although there is a large standard deviation, the mean time to complete the exam was about 53 minutes; which was perhaps a little lengthy. The target time for completion of each of the instruments was 40 min, it was theorized that any assignment that took significantly longer would be counterproductive as it would affect the students' motivation and interest.

The large standard deviation is most likely due to differing levels of motivation due to the fact that the bonus points were only given for completing the test and not based on performance. Motivation may also have been affected because the bonus for completing the assignment was so small, just 1% of the final grade.

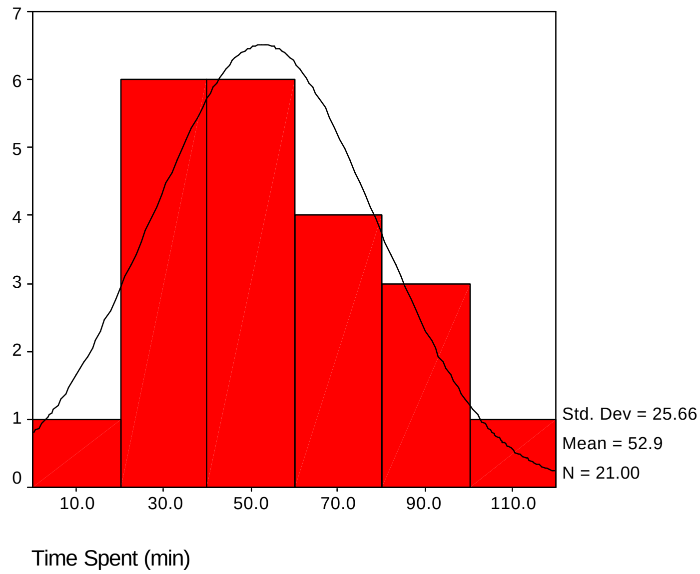


Figure 5.2 Histogram of Time Taken to Complete Preliminary and Pretest Questions

Scores for the take home test were broken down into scores on preliminary questions and scores on pretest questions (please note distinction above). Scores on preliminary questions are presented in histogram form in Figure 5.3; the scores were out of a possible 34 points. The distribution is somewhat normal with a mean of 19 points (55.8%) and a standard deviation of 7 points (20.8%). A broad range of student abilities is clearly present, revealing that different students come into the truss portion of Statics with different skills and knowledge bases. This spread of scores allows for comparison across groups in the future to assure that study populations are similar.

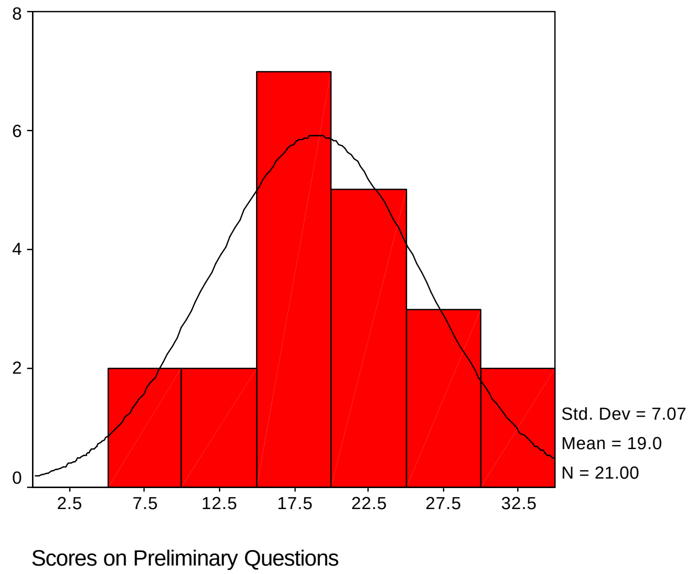


Figure 5.3 Histogram of Scores on Preliminary Questions

The scores on the pretest questions are presented in Figure 5.4 below. The distribution once again is somewhat normal though slightly heavy on the lower scores. This skewing is expected since students have been exposed to little truss information. With a mean of 10 (31.3%) and a standard deviation of 6.6 (20.7%) there is definitely room for improvement of truss knowledge. Most of the points that students did receive on these questions, 43.7% of the total, came from qualitative questions about trusses (e.g. Members of a truss are assumed to be connected by smooth pins, true or false) which only accounted for 19% of the total points possible (6 out of 32). The fact that most of the points students earned on the exam came from one small set of questions can probably be explained by the fact that the information required to answer these questions was presented to the students in the one lecture on trusses that they had before taking the test. Guessing on true/false questions could be another explanation for higher scores on the qualitative questions; when nothing is known about the material being tested, it is easier to circle T or F than to write down something on an opened ended question or calculation.

As mentioned previously, the posttest was not administered in its intended form. As one section of class only took half the test, and the other section took the other half, the results cannot be combined for comparison to the pretest and as such are not presented here. Considering the purpose of the formative study was to evaluate the instruments, this is not an issue because the posttest was identical to the pretest and thus proper evaluation of the pretest applies to the posttest as well.

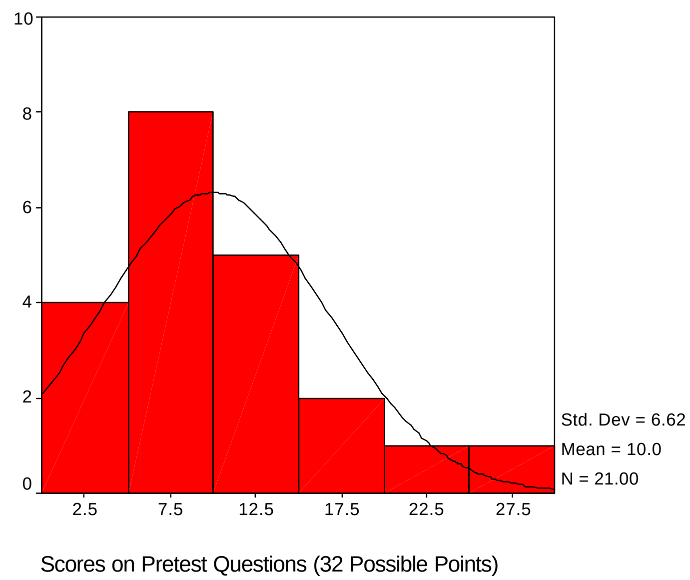


Figure 5.4 Histogram of Scores on Pretest Questions

The distribution of students' scores on the exam question is pictured in Figure 5.5. The distribution is clearly bimodal which is interesting considering that scores on the other assessment instruments tended to be somewhat normal. The course instructor, however, as well as other faculty members who teach this course did not find this unusual, pointing out that truss analysis is a fairly straightforward process that students either do or do not know how to do. One of the professors referred to truss analysis as "light switch problems" noting that students are either on or off. As can be seen in the figure, almost half of the students did not even earn half of the credit available for this problem. Clearly there is some room for improvement and it was agreed among the statics instructors that perhaps an intervention involving educational technology could help these students who are otherwise confused by truss analysis.

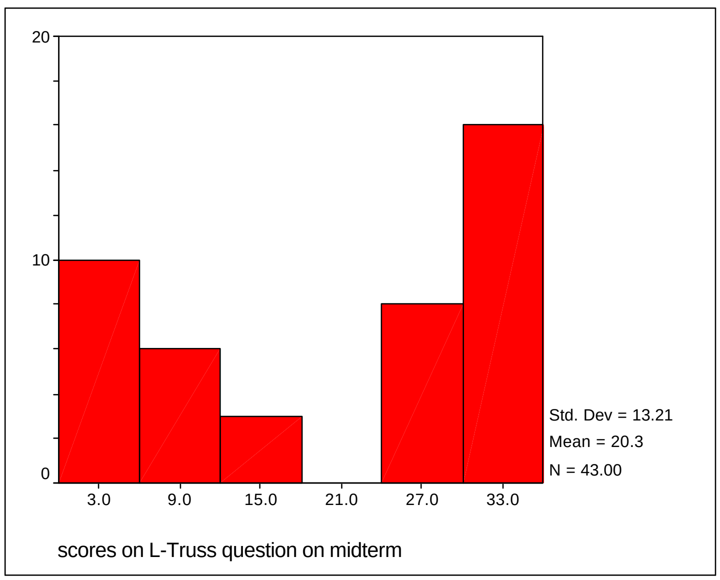


Figure 5.5 Histogram of Scores on Exam Question (35 Possible Points)

These results were analyzed to assess the usability, reliability and validity of the testing instruments.

5.4.1 Usability

A few usability issues were brought up and corrected during the formative study. Most of the corrections involved revising question wording, adding questions, or changing the format of the instruments.

One general revision that was that students would no longer be encouraged to answer *I don't know* to problems they were not able to complete on future assessments. It was observed from these formative results that, despite asking students to try to complete the problem before answering *I don't know,* many of them used this option as an excuse to avoid some of the more difficult questions. It was assumed students would put more effort into each of the problems on the assessment instruments if this option was removed. Observations from final assessments support this assumption.

5.4.1.1 Preliminary Questions

Some concerns about the preliminary test became apparent when students filled it out. A number of problems were consistently missed, which is acceptable if the problem is designed to be challenging. One such problem asked the students to solve for the reactions of a Howe truss placed and loaded on an incline. Students have studied equilibrium and should be able to solve for reactions but the problem appears intimidating and is challenging and thus is not expected to be completely solved by many of the students. However, many students missed other problems that they should have been able to answer with ease. This poor performance led to revisions of two questions. The first question initially read as follows:

How many reaction forces do the following types of supports provide?

Roller _____

Pin _____

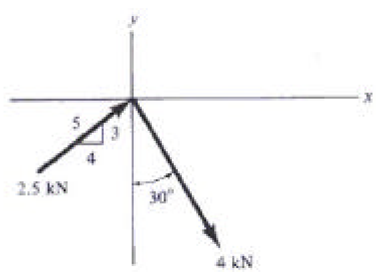
Rocker _____

Fixed _____

In a meeting with faculty members, it was agreed that the reason that many students missed this question may have been the use of the phrase *reaction forces* because any force could be broken down into any number of component forces (for example a rocker provides only one reaction, but when placed at an angle is sometimes broken down into Cartesian components for ease and thus may be confused as having two reaction forces). To alleviate this concern, the wording of this question was revised for future use to read: *How many **unknowns** are associated with the following types of supports?*

The second problem to be revised initially read as follows:

Resolve the following force vectors into their x and y components. Add the vectors and determine the magnitude and direction, Q measured from the positive x-axis, of the resultant force.



Many of the students missed points on this problem because they did not complete each of the steps required in the problem statement. To correct this concern, the problem has been revised by breaking it down into a four separate problems where each separate problem will require just one step of the original problem. In addition to these revisions, faculty members requested that a question on cross-products be added to the preliminary question set as this is another skill students could utilize in truss analysis; a simple question on the topic was added for future use.

5.4.1.2 Pretest/Postest

A few revisions were made on the pretest/posttest as well. One set of questions that were revised related to a truss with unknown dimensions. Load directions were given but the magnitude of the loads were not. Based on this minimal, yet sufficient, amount of information, students were asked to determine whether specific members were in tension, in compression, or zero force members. This set of questions was considered important by the group of faculty members who teach the course because it asks students to think qualitatively about the truss as a whole rather than crunching numbers around a joint or section. Student responses, however, revealed that many did not take these problems very seriously. The mean percentage of number of problems correct on this set of problems is 0.95 out of 3 (31.7%). This is understandable before students had learned trusses, but on a posttest given to students a week later, after they had completely covered trusses, revealed only a modest change in this result, 1.14 out of 3 (38%). Few of the students' returned tests showed any sort of scratch work, sketches, or notes on these problems revealing that they probably did not spend much effort on these problems and may have even simply guessed on them. To alleviate this problem, the revised questions asked students to not only identify the member type (i.e. compression, tension, or zero-force member) but to explain their choice in short answer form.

In addition to this problem set, another group of questions was added that asked students to think qualitatively about trusses. For this set of questions, students were given two identical trusses, both of which with minimal yet sufficient information (i.e. no dimensions or quantitative values for loads). The two trusses differed, however, in that the first truss had a single load applied in the center of the truss and the second truss had the load distributed across the top chord of the truss (see the final form of the pretest/posttest in Appendix C). Students were asked how the difference in loading affected the internal loads of various members of the truss and why. Student responses in short answer form provided a method to qualitatively assess students' truss knowledge.

An additional, simple, quantitative analysis problem was added and the K-truss analysis problem was taken out and replaced by the L-truss problem that was used on the midterm (figure 5.1) due to the fact that a similar K-truss problem was going to be explicitly exemplified in future classes. Thus, the final form of the pretest/posttest (see Appendix C) contained two analysis problems, one which lent itself to solving with method of joints and the other which lent itself to solving with method of sections, two qualitative analysis problems mentioned in the preceding paragraphs, and a set of qualitative questions about the nature and assumptions of trusses in general.

5.4.1.3 Exam Questions

The exam question was not revised but was incorporated into the pretest/posttest as mentioned above. Future exam questions were not designed during the formative study because the instructors of the courses to be used in the study had not been chosen and it was assumed that they would want to have some input on the questions they would be giving on the exam. As such, exam questions were designed later and are mentioned at another point in this document.

5.4.2 Reliability

Accepted statistical methods were employed to assure the reliability of the preliminary test and the pretest/posttest.

For the preliminary questions, the results from both sections were combined. Twenty-one students completed the preliminary questions and as such there were 21 subjects. Because reliability of a single instrument containing multiple items was desired, as opposed to reliability across instruments or across time, the statistical method chosen for this measurement was the Split-Half reliability measure (Whitley, 1996). To complete this statistical test, the preliminary questions were divided into two parts. The responses to these split parts were then compared, via the Guttman Split-half method, as if they were two different sets of questions (SPSS, 1999). The method returned a reliability coefficient of 0.772 which exceeds the recommended minimum coefficient of 0.7 (Whiley, 1996). The same method was used to test the reliability of the pretest/posttest questions and yielded a coefficient of 0.835 which also exceeds the minimum value. Thus, it can be inferred the instruments are sufficiently reliable.

5.4.3 Validity

All of the instruments were presented to a group of faculty members who teach statics for their approval. All of the faculty members agreed that the instruments tested each of the concepts and topics that a student should know either before or after learning truss analysis. This type of agreement between experts validates the instruments by lending content-related evidence of validity as mentioned in Whitley (1995).

Whitley (1996) also lists criterion-related evidence as another way to validate an instrument. Essentially, criterion-related evidence is the degree to which a measure is related to some other measure or criterion. In this case, in order to validate the preliminary questions, the results of the preliminary questions were compared to the students final grades and their grade on the truss problem of the midterm exam. Because final grades are ordinal data, a nonparametric method was employed. The Spearman rank order comparison was made between the three measures with a Bonferroni adjustment for two comparisons. Both correlations were significant at the 0.05 (0.025 after the Bonferroni adjustment) with the correlation between the preliminary results and the final grade yielding a coefficient of ?=0.538 ($p<0.025$) and the correlation between the preliminary results and the grade on the midterm truss analysis question yielding a coefficient of ?=0.548 ($p<0.025$). These tests lend further evidence to the validity of the preliminary questions. The posttest was also compared to the exam question and revealed a highly significant relationship (?=0.891, $p=0.00002$).

5.5 Summary and Conclusions

The formative study was successful at accomplishing the objective of evaluating the instruments that would later be used in the actual study. The formative study led to a few minor revisions on the preliminary test and some major revisions on the pretest and posttest. As the preliminary test results were not only fairly well distributed but also significantly correlated to both final grade and ability to analyze a truss on an exam, it is concluded that it will act as an appropriate tool for comparing the abilities of students in future sections to assure that the various sections are starting at the same level. Similarly the pretest/posttest was significantly correlated to the students' ability to analyze a truss on an exam and as such are appropriately measuring what they were designed to measure. Furthermore, the preliminary test and the pretest/posttest proved to be reliable measures which suggests that they can be used effectively in the future with a different study population.

CHAPTER 6

INTERVENTION IMPLEMENTATION

As mentioned in chapter 3, the intervention was not implemented as proposed. The fundamental research design did not change, a quasi-experimental nonequivalent comparison group design was still used, however the implementation of that design did differ from the proposed intervention in a number of ways. This chapter will begin by explaining the changes that took place and will conclude by describing precisely what occurred during the interventions.

6.1 Changes to Proposed Intervention

Most of the changes that took place fell into one of two categories: student centered changes and instructor centered changes. Each of these changes will be detailed here as well as why the changes were made.

6.1.1 Student Centered Changes

Student centered changes were made based on information gathered during the formative stage of this research.

6.1.1.1 Student Population Information

The formative study, described in the previous chapter, proved to be valuable not only in that it was a means for successfully evaluating the reliability and validity of assessment instruments, but it also provided important insight into the nature of the subject population. Prior to the formative assessment, it was assumed that a majority of the population were second-year civil engineering (CE) students. In the course of the formative study, however, it was discovered that CE students were a minority in CEE 2020, and that most of the students were either industrial engineering (IE) or electrical engineering (EE) majors. Beyond this, it was determined that most of the IE and EE students were juniors or seniors and that a number of them were enrolled in their final semester and were graduating soon.

Based on this information, two years of previous course records for CEE 2020 were researched. The results of this research into the records are given in table 6.1, which presents the number of students in each section who graduated within six months of CEE 2020, and table 6.2, which presents the breakdown of each section by major.

This investigation revealed that the population trends discovered during the summer of 2002 were not isolated, and that all of the sections over the past two years followed the same pattern. More importantly, it was discovered that as many as forty percent of the students in some sections (Spring 2001, section I) of the course had graduated within six months.

This population information was important for a few reasons. First, the fact that many of the students who would participate in the intervention would no longer be on campus in six months was of great concern because, as proposed, the long-term retention test was to be administered twenty-five weeks (approximately six months) after the intervention. It was assumed that students would be less likely to participate in follow up studies after graduating. Furthermore, students who have graduated could no longer be contacted through their school-administered email account, as it would no longer be active, nor could they be tracked through student records.

Table 6.1 Number of Students per Section of CEE 2020 Who Graduated Within Six Months.

		Graduated Within Six Months				
		Yes		No		
	Section	Count	Percent	Count	Percent	Total
Fall 1999	B	1	4.8%	20	95.2%	21
	C	0	0.0%	50	100.0%	50
	D	3	6.1%	46	93.9%	49
	E	3	5.9%	48	94.1%	51
	F	0	0.0%	39	100.0%	39
	G	0	0.0%	17	100.0%	17
	I	2	4.8%	40	95.2%	42
		Yes		No		
	Section	Count	Percent	Count	Percent	Total
Spring 2000	B	1	2.5%	39	97.5%	40
	C	9	20.9%	34	79.1%	43
	D	1	4.2%	23	95.8%	24
	E	10	24.4%	31	75.6%	41
	F	2	16.7%	10	83.3%	12
	G	0	0.0%	7	100.0%	7
	I	3	7.5%	37	92.5%	40
	J	0	0.0%	28	100.0%	28
		Yes		No		
	Section	Count	Percent	Count	Percent	Total
Fall 2000	B	2	4.1%	47	95.9%	49
	C	0	0.0%	38	100.0%	38
	D	0	0.0%	20	100.0%	20
	E	0	0.0%	35	100.0%	35
	F	0	0.0%	14	100.0%	14
	G	0	0.0%	27	100.0%	27
	H	0	0.0%	66	100.0%	66
	I	2	4.3%	44	95.7%	46
		Yes		No		
	Section	Count	Percent	Count	Percent	Total
Spring 2001	B	16	33.3%	32	66.7%	48
	C	5	12.2%	36	87.8%	41
	D	4	8.5%	43	91.5%	47
	E	2	5.6%	34	94.4%	36
	F	0	0.0%	17	100.0%	17
	G	5	11.1%	40	88.9%	45
	I	21	40.4%	31	59.6%	52
		Yes		No		
	Section	Count	Percent	Count	Percent	Total
Fall 2001	B	1	9.1%	10	90.9%	11
	C	0	0.0%	39	100.0%	39
	D	1	3.8%	25	96.2%	26
	E	0	0.0%	41	100.0%	41
	F	4	8.9%	41	91.1%	45
	G	3	6.7%	42	93.3%	45
	H	9	12.2%	65	87.8%	74
	I	4	14.8%	23	85.2%	27

Table 6.2 Breakdown of Population by Major per Section of CEE 2020.

Semester	Section	CE	IE	EE	ChE	Other	Total
Fall 1999	B	10	1	2	4	4	21
	C	7	33	1	5	4	50
	D	15	18	1	12	3	49
	E	6	31	3	4	7	51
	F	2	21	0	8	8	39
	G	1	6	2	7	1	17
	I	6	20	1	10	5	42
Spring 2000	B	16	12	1	5	6	40
	C	8	26	3	5	1	43
	D	6	4	4	1	9	24
	E	7	23	2	6	3	41
	F	0	8	2	1	1	12
	G	0	2	2	0	3	7
	I	5	25	1	4	5	40
	J	5	12	3	3	5	28
Fall 2000	B	10	26	6	2	5	49
	C	10	19	2	3	4	38
	D	4	4	2	2	8	20
	E	7	21	4	2	1	35
	F	4	4	0	6	0	14
	G	6	7	9	2	3	27
	H	25	30	6	2	3	66
	I	7	19	12	6	2	46
Spring 2001	B	5	30	7	3	3	48
	C	11	20	1	2	7	41
	D	15	5	17	6	4	47
	E	7	16	4	6	3	36
	F	3	5	3	3	3	17
	G	6	19	11	3	6	45
	I	3	23	14	5	7	52
Fall 2001	B	5	0	3	3	0	11
	C	8	23	5	1	2	39
	D	7	6	9	2	2	26
	E	12	21	5	0	3	41
	F	1	19	22	2	1	45
	G	5	20	16	3	1	45
	H	29	28	14	1	2	74
	I	5	9	8	2	3	27

The second reason was also related to the long-term retention test. As proposed, the test would be completed voluntarily by students who were sent the test via email or would be completed in a follow-up course to CEE 2020. In the interim between project proposal and implementation, concerns were raised about the successfulness of emailed tests. It was then assumed that many students would not be sufficiently motivated to voluntarily complete and return a test that was emailed to them (this assumption proved to be very true as will be seen later in this chapter). It was decided that some form of motivation would have to be provided to students in order to encourage them to complete the long-term retention test. It was also decided that the best way to contact students and to motivate them to participate in the long-term study would be to incorporate the test into a follow-up course. The population data gathered during the formative evaluation was of concern because no follow-up course consisted of a population similar to CEE 2020. Specifically, no course exists that primarily consists of sophomore CE students, and junior and senior IE and EE students who had just completed CEE 2020.

Third, the formative data revealed that the research population was not homogeneous. Specifically, based on the data it could not be assumed that senior-level IE students would have a similar background and skill level to sophomore-level CE students. IE students have been in school longer and perhaps may have a more refined set of problem-solving skills than the more inexperienced CE students. Also, IE students take a whole different set of courses than do CE students. This concern was mentioned in a previous section of this document on variables but it is mentioned here again because it influenced a change that took place in the intervention. It is important to note that differences *within* subject groups does not preclude comparisons *between* subject groups, it simply means that the groups must be evaluated to ensure that the population distributions are similar in the areas of major and rank before other comparisons could be made or that the results must be broken down by major and rank.

6.1.1.2 Changes Based on Population Data

Based on the population data gathered during the formative stage of the research and the concerns mentioned above that arose from these data, a couple of student centered changes were made to the proposed intervention. The first change that took place involved the dissemination and completion of the long-term retention test. As mentioned previously, it was decided that the best way to collect the long-term retention data would be in a follow up course. This would be done with the assistance of an instructor in one of a few courses that students take after CEE 2020 who would agree to hand out the test in class and give students credit for returning the completed assignment.

An obvious consequence was that only CE students would be reached through this method and long-term data would not be gathered from EE and IE. This is, however, an acceptable consequence based on the third concern mentioned in the previous section that IE and EE students differ considerable from CE students and that these differences would only be accentuated by the courses taken after CEE 2020. These differences would require the long-term results to be broken down by major, with each major studied individually. Rather than look at three majors individually, it was decided that the complete attention of the long-term research would be focused on CE students only. Other factors in this decision were the assumptions that instructors in IE and EE follow up courses may not be willing to participate in a study run by unfamiliar researchers (as opposed to CE instructors, many of whom are familiar with the researchers) and that a good number of IE and EE students would have graduated within six months and would not be available for a follow up exam. Learning and retention data would still be gathered from all students in CEE 2020, but long-term retention data would only be gathered from CE students.

The second student centered change is a natural extension of the first. Because only CE students would be assessed long-term and because CE students were discovered to be a minority in most sections CEE 2020, it was decided that the three sections that would be chosen as the sample populations would be the three sections with the most CE students in them so as to maximize the number of subjects participating in the long-term retention test. Because a quasi-experimental research design had been chosen this non-random selection of experimental groups did not affect the efficacy of the design. The only implication this decision did have is that it precluded the selection of groups based on the matching of non-experimental variables such as gender and ethnicity. As mentioned previously, however, there was no reason to assume that these variables would have any appreciable affect on the dependent variables and as such it was less important to control for these variables than it was to control for the number of CE students. In the analysis of the data, though, comparisons would be made to assure that this assumption was true; that is to say that gender and ethnicity would not affect the dependent variables.

In summary, the two major student centered changes were as follows. First, the long-term retention test would be administered in one of a few CEE courses that are follow-up courses to CEE 2020 and would only be administered to CE students. Second, the sections of CEE 2020 that would be chosen to participate in the experiment would be the sections containing the greatest number of CE students (pending instructor participation of course). These changes were proposed to the research committee and were approved prior to being implemented.

6.1.2 Instructor Centered Changes

Upon choosing the three sections of CEE 2020 that would serve as the experimental and comparison groups for the study, the instructors of those sections raised some concerns about the intervention that let to changes in the intervention implementation.

6.1.2.1 Special Session

The biggest concerns that instructors had concerning the intervention in their specific classes were the details of the intervention: specifically how, where, and when the intervention would actually take place. As proposed, the intervention would take place during a special session, like a recitation session, outside of normal classroom hours. There are, however, no recitation sessions officially scheduled for any sections of CEE 2020 and instructors were extremely hesitant to require students to participate in a class session that was not officially scheduled (i.e. listed in the Institute's schedule of courses). The instructors were also concerned about taking class time away from a course that was already on a very tight schedule and thus would have preferred that the student participation be on a voluntary basis. The researchers feared that this would compromise the design and results of the research and were thus hesitant to agree to such a stipulation.

After a great deal of discussion, a compromise was made. The special intervention session, the session during which the independent variable would be introduced, would take place during one regularly scheduled lecture period. This was a compromise because it took some time away from the instructors but it also allowed the researchers to require students to be present as it is within the instructors authority to require attendance of regularly scheduled class sessions. An unfortunate consequence of this compromise was that it limited the intervention to only an hour as opposed a three hour special session as originally designed. This was of concern to the researchers but it was all that was allowed. This was the only situation to which both the instructors and the researchers agreed and so it is the plan that was later implemented. More specific details and a precise schedule of events relating to the special session will be given later in this chapter.

6.1.2.2 Completion of Assessment Instruments

The instructors of the specific sections in which the intervention would take place also had concerns about taking time away from class to complete the assessment instruments. The instructors' schedules for CEE 2020 were extremely rigid and did not allow for lecture time to be used for other activities. As such, a number of changes had to be made to the assessment instruments and how they were disseminated and completed. Recall that the proposed project called for four different assessment instruments: a pretest, posttest, retention posttest, and long-term retention posttest. The impact of the instructors concerns on each of these instruments will be discussed.

Motivation

As will be explained shortly, instructors required that some of the assessment instruments were to be completed by students at home. In order to motivate students to complete the assignments at home and to give them their best effort, all the involved instructors agreed to adjust their grading scheme and set aside five percent of the students final grades to be awarded for participating in the research project. Participation, as explained to the students, included completing a few special take home assignments to the best of their abilities. One of these take home assignments was the pretest.

<u>Pretest</u>

As explained in chapter 5, the pretest was designed to include two different types of questions, preliminary questions and truss questions. The preliminary questions were concerned with information that students must know in order to solve truss problems and included topics such as trigonometry and equilibrium. The truss questions dealt with trusses both qualitatively and quantitatively; these questions also were to appear in identical form on the subsequent instruments (i.e. the posttest, retention test, and long-term retention test). The two types of questions were shuffled and the resulting pretest was ten pages in length.

Meetings with the instructors revealed that there was no class time in which to administer the pretest and that it would have to be a take home assignment. Provided that the instructors were willing to offer the aforementioned motivation, the researchers were willing to agree to a take-home pretest.

Instructors, however, were concerned that students would be overwhelmed by an ten page assignment and would, perhaps, not give it their best effort. They suggested cutting out the truss questions and just asking the preliminary questions on the pretest. A consequence of this decision would be that the learning variable would no longer be a differential variable (variable measuring the difference between performance before and after the intervention) but would simply be a direct measurement of performance on the truss questions on the posttest. This was a consequence that the researchers were willing to accept in return for increased validity on the preliminary questions. The pretest, as a result, was trimmed to only included preliminary questions, the results of which would be used to compare the experimental groups to assure that previous knowledge was similar among all groups. The final form of the pretest is presented in Appendix D.

Posttest

The instructors were satisfied with the content and length of the posttest as designed (see chapter 5) and thus the posttest remained unchanged. The instructors did allow forty minutes worth of class time for the completion of the posttest.

Retention Test

Again, the instructors were satisfied with the form and content of the retention test, which was identical to the posttest, and as such it remained unchanged. The retention test was designed to be administered approximately ten weeks after the intervention which ended up being the last week of classes. The instructors, who from experience knew that this would be a busy time for them and that they would be both catching up and wrapping up the course, did not foresee having time to allow for the completion of the retention test during class. The instructors requested that the retention test, as with the pretest, be assigned as a take home assignment. The researchers agreed with the request provided that the instructors follow through with the promised motivation as mentioned above.

Long-term Retention Test

The instructors of the classes in which the intervention would take place had no opinions or concerns related to the administration of the long-term retention test because the students being studied would not be in their classes twenty-five weeks after the intervention. The changes to the administration of the pretest were mainly student-centered and were detailed earlier in this chapter.

6.1.2.3 Remarks

As described above, a number of major methodological changes were instigated based on the input from the instructors of the courses in which the intervention would be implemented. There are two important remarks that must me made concerning this matter. First, note that these are the only circumstances under which these instructors would allow the research to take place in their sections of CEE 2020; so while some of these concessions may seem to be detrimental to the integrity of the research, in actuality these concessions were required in order to allow the research to occur. Second, the instructors were very excited about the research process and willingly gave up personal time, class time and control of some class elements, including exams, in order to see that this project could take place and succeed. Despite their concerns and objections, this project could not have occurred without their help and assistance.

6.1.3 Summary of Changes

Based on information gathered from CEE 2020 student records and input from instructors, a number of changes were made in the research process. These changes are summarized in bulleted form below.

- Long-term retention data was only gathered from CE students.
- Long-term retention data would primarily be gathered in a CEE course subsequent to CEE 2020.

- Sections of CEE 2020 with the most CE students were chosen as the experimental and comparison groups for the study.
- Five percent of the students' final grades in the chosen sections were based on participating in the research project.
- The pretest included only preliminary questions.
- The pretest would be a take-home assignment.
- The special session, wherein the dependent variable would be introduced, would take place during one regularly-scheduled class session.
- The learning or performance variable would no longer be a differential variable measuring the difference between results on truss questions before and after the intervention.
- The learning or performance variable would simply measure the performance of students on a posttest containing truss questions after the intervention.
- The retention test would be administered as a take-home assignment.

The above changes were proposed to the research committee in August of 2002 prior to implementation and approved. These changes did alter the research process significantly enough to require a new review by IRB and the research continued under exempt status.

At this point, the proposal and planning stages of the project were completed. The software selection process had successfully yielded two very good software titles. A formative study was conducted which revealed some important information regarding the population which would be studied in the actual intervention. The formative study also served to refine, validate, and measure the reliability of assessment measures. The research process was refined through the input and assistance of statics instructors who would be instrumental in carrying out the project in their own courses and the research committee was informed and up to date on the new research methodology. All preparations were made and plans set for the research intervention to actually take place. The implementation of this intervention as it took place is described in the following section.

6.2 Intervention Process

As mentioned previously, this study was replicated in order to minimize the possibility of a selection bias that might be present in a quasi-experimental research design. The study, then, was completed twice for this project, once with students who took CEE 2020 in the fall of 2002 and once with students who took CEE 2020 in the spring of 2003. These individual studies will be discussed separately below.

6.2.1 First Implementation, Fall 2002

The first implementation of the study began in the fall of 2002 and continued through the spring of 2003 when the long-term retention results were collected. This section will describe the process of that study.

6.2.1.1 Subject Selection

In the fall of 2002, eight sections of CEE 2020 were taught by six different instructors, two of which taught two sections of the course. Upon the close of late registration, one week after courses began, data were gathered on each of the sections to determine which sections of the course had the highest number of CE students enrolled in them. These data are presented in table 6.3, which shows the number CE students and the number of total students for each section of the course taught in the fall of 2002. Sections C, E, and H were chosen to be the three different sample groups for the study because they contained the greatest number of CE students.

Table 6.3 Number of CEE Students in Each Section of CEE 2020, Fall 2002.

Section	CEE Students	Total Students
B	8	21
C	17	45
D	3	35
E	14	43
F	5	47
G	3	45
H	29	71
I	3	22

Sections C and E were taught by the same instructor, referred to herein as Instructor A, and section H was taught by Instructor B. The reasons for higher enrollments of CE students in these sections were not coincidental. Sections C and E were limited during initial registration to sophomores only and, as mentioned previously, most of the IE and EE students who take the course are juniors or seniors and were thus prohibited from enrolling during early registration without permission from Instructor A. Instructor B has an excellent teaching record and is a highly sought after instructor for CEE 2020. Because of this, section H fills up quickly during early registration with IE and EE students again because most of them are juniors and seniors and are allowed to register for classes before sophomores. Instructor B, however, usually opens up the course for more students or overloads many students (both methods allow more students to register for the class than the maximum number listed in the official institute course listings) later on during registration which allows many CE students to enroll in this section. As a result of this, section H usually has about fifty percent more students (70 on average) enrolled than was initially allowed during early registration (45) and around seventy-five percent more students than the average maximum number of students per section (40). Clearly there is a high probability of selection bias which lends evidence to the need for replication of this study, which replication was completed and is described later. Instructors A and B agreed to allow the study to take place in their sections given that the concessions mentioned earlier in this chapter were met.

Section C met on Mondays, Wednesdays and Fridays from 10:00 A.M. to 11:00 A.M. and was assigned to the Content subject group, or the group that would use the content-type software during the special session of the intervention. Section E met on Mondays, Wednesdays, and Fridays from 12:00 P.M. to 1:00 P.M. and was assigned to the Tool subject group, or the group that would use the tool-type software during the special session of the intervention. Section H met on Tuesdays and Thursdays from 8:00 A.M. to 9:30 A.M. and was assigned to the comparison or control group, the group that would solve problems during the special session without the assistance of any software.

While these assignments were made for no particular reason, they were not randomly assigned either. That is to say, that a specific process (rolling of a die or using a random number table for example) for assigning the groups was not developed and followed, assignments were simply completed at the discretion of the researcher who made the assignments for no specific reasons. This is not of concern because the nonequivalent comparison group design does not require random selection or assignment and because replication will reduce the possibility of an effect resulting from a selection bias.

More detailed information about the sample groups will be presented in the results chapter, which will describe various demographic data about the groups and make comparisons between the groups based on these demographics.

6.2.1.2 Pretest

Prior to the intervention, students were asked to complete a pretest to gauge their prior knowledge, the results of which would be used to show equality between the groups. As previously mentioned, Instructors A and B did not have sufficient class time available to allow the pretest to be completed in class and thus required that the pretest be given as a take home assignment. Motivation to complete this take home, as well as other portions of the research project, was provided by the instructor in the form of a percentage of the students' final grades. Both Instructor A and B informed students on the first day of classes, and on the written syllabus, that five percent of the their final grades would be awarded based upon their individual participation in a research project that would be conducted throughout the duration of the course.

The instructors did not want the students grades, however, to be based on assessment instruments that they did not write nor did they want the grades to be based on information that was not directly taught in class but would be assessed in this research project (i.e. prior knowledge such as linear algebra). Because of this, they required that the five percent would be based simply on participation and not on performance. Thus, there was motivation for students to participate but not necessarily to perform to the best of their abilities.

To provide motivation for performance, a cover page was attached to the pretest, which described the benefits of participation and asked the students to give the assignment their honest effort so that an accurate assessment could be made. The cover page was written with the assistance of the institute's Director of Assessment and is shown in figure 6.1. The document explains that this assignment is part of a research project and then describes, in a very general sense, the purpose of the research and that it will have an impact on important decisions involving engineering education at Georgia Tech. The intent was to give students a proactive role in their educational experience and encourage them to take action by completing the assignment. The cover page goes on to describe what will be required of them and what they will receive in return and then sincerely asks for the students to give an honest effort. Confidentiality is assured followed by some concluding instructions on completing the assignment. Additional motivation for performance was contributed by the instructors who informed students that although the assignment would not be graded, it was very important to them (the instructors) and that they wanted the students to do their best.

PLEASE READ THESE INSTRUCTIONS IN THEIR ENTIRETY AND SIGN YOUR NAME TO VERIFY THAT YOU HAVE DONE SO.

In order to improve instructional methods in the College of Engineering and thus make Georgia Tech engineers even more knowledgeable and competitive, a research project is being conducted in a number of different sections of CEE2020.

As engineering education continues to evolve and change, it is important for educators to understand the nature of how students learn and use this understanding in our instructional methods. The results of this and other assignments will provide insights to the College of Engineering, the Office of Assessment, the Associate Provost for Institutional Development and others who are responsible for making important decisions regarding instructional, curriculum, and program development.

Your participation is a required part of the course and is worth 5% of your final grade. To earn these points, you must complete three assignments, the first of which is attached as a take-home assignment. The second assignment will be done in class and the third will be given as a take-home assignment later in the semester. This assignment should take about 25-40 minutes to complete. Your honest effort to answer each question correctly will provide us with accurate results that will be used to improve the educational experiences of students at Georgia Tech.

Your instructor will be the only person who will have access to your individual results. You may be assured of complete confidentiality; aggregate data only, with the names removed, will be published to offices on campus or off.

This assignment is to be performed by you without the help, in any fashion, of any other person. The use of your textbook, course notes, or any resource other than a calculator is not permitted. By signing your name below you acknowledge that you have read and agree to the above conditions.

Signature Date

Printed Name

Figure 6.3 Pretest Cover Page

Though there is no quantitative evidence that these measures actually had any effect on performance motivation, anecdotal evidence in the form of researcher observations revealed that most students did put forth a substantial amount of effort. There were also stark contrasts between the returned assignments in the fall and those returned in the summer when students were encouraged to write *I don't know*, a choice which was opted for by many in the summer sections; the students in the fall completed more problems than their summer counterparts. While it cannot be assumed that all students put forth their best effort on each problem, there is no reason to assume that the degree of effort varied between the three fall sections participating in the study. As such, effort was not measured but was assumed to be equal among all research groups and motivation to perform will not be further discussed.

The final form of the pretest is shown in Appendix D. The pretest was given to students on the first day of the intervention and they were required to return it on the day of the special session. This gave students five days to complete the assignment at home. Students were given verbal instructions by the researcher similar to the instructions on the cover page and the assignment was self-explanatory. Most of the students completed and returned the assignment with almost no questions or concerns. Though the cover page informed students that the test should take 25-40 minutes to complete, students were allowed to take as much time as possible but they were asked to record the times at which they began and completed the assignment. The results of the pretest, including the times taken to complete the assignment, will be discussed in the following chapter.

6.2.1.3 Instruction

Sections C, E, and H received almost identical instruction on truss topics during the intervention. Instructors A and B agreed to allow the researcher, who was teaching another section of CEE 2020 at the time, to teach their sections during the intervention. As such, all three sections had the same guest lecturer, learned the same principles and skills, and worked through the same example problems. Because truss analysis is a fairly straightforward topic, similar questions arose in each of the sections and no topic or concern was addressed in one section that was not also given consideration in the other sections as well. In short, the researcher lectured very carefully so that each section would be working with the same knowledge base and skill set, in an effort to assure that no section had an advantage over the others.

Each of the instructors agreed to give up approximately three hours and forty minutes of class time for the intervention, which included two hours of traditional lectures on trusses that were conducted by the researcher. The first hour of the intervention was spent teaching one of two truss analysis methods: the method of joints. A few minutes were taken from this first hour to briefly and in very generic terms (so as not to bias the participants) explain the research project and hand out the pretest. The first hour took one whole lecture period for sections C and E, and two thirds of a lecture period for section H (because this section meets for an hour and a half twice a week as opposed to one hour three times a week). The remainder of this first lecture period for section H was spent on the method of sections, the second truss analysis method. An hour was devoted to this method in each of the sections so one third of the following lecture period for section H was devoted to this method. The second hour of lectures took one whole lecture period for sections C and E. Ordinarily, this is where both the instructors would conclude their instruction on trusses and would move to a new topic. The intervention, however, continued with the inclusion of a special session and a posttest.

The third hour of the intervention was devoted to the special session and the final forty minutes were devoted to the posttest, both of which will be discussed in following sections. The special session took the third whole lecture period for sections C and E and completed the final hour of the second lecture period for section H. The final forty minutes of the intervention took place during the first forty minutes of third and forth lecture periods for sections H and C and E respectively, the remainder of these lecture periods were given back to the instructors who moved onto a new topic at which time the intervention was complete. Figure 6.2 shows the timeline of events during the intervention in relation to the scheduled lecture periods for the different sections.

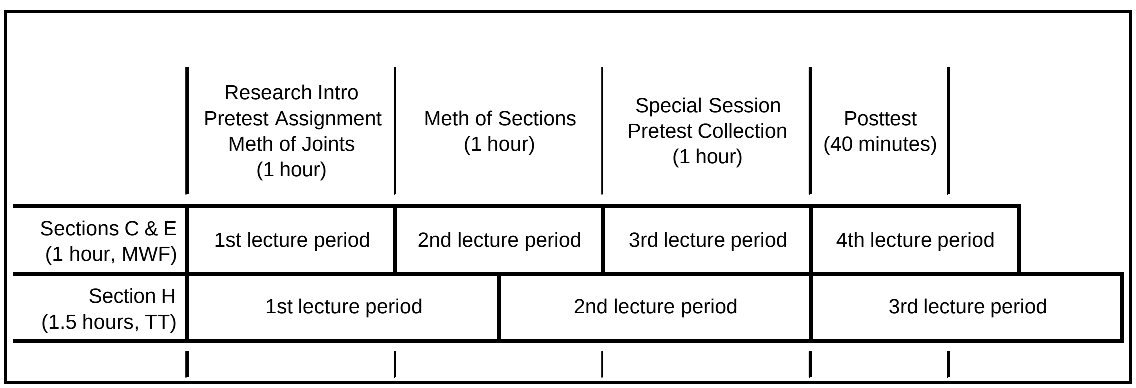


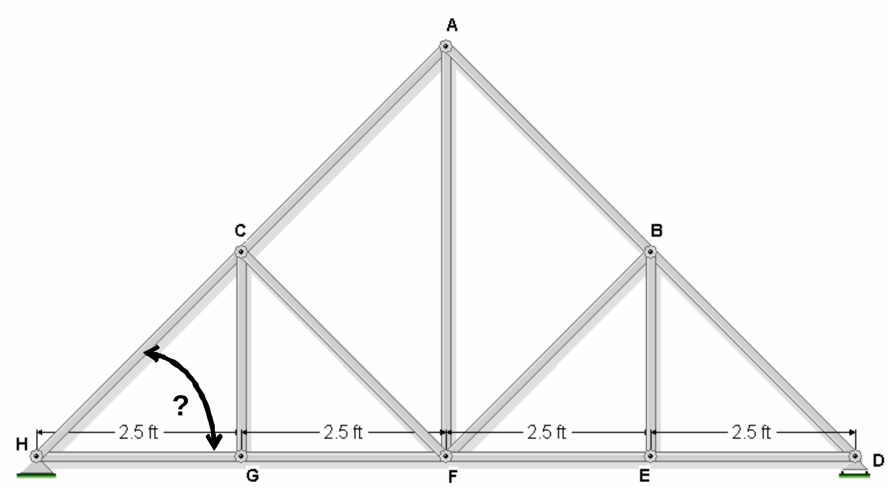
Figure 6.2 Timeline of Intervention, Fall 2002

6.2.1.4 Special Session

There was much discussion among the research committee about what would exactly be done during the special session. When the instructors limited the time of the special session to just one hour, the options became fewer and the discussion was soon settled. The primary concern was equality or what type of variable should be equalized among the groups. Those using a piece of software could probably finish problems much more quickly than those solving problems by hand, especially considering that the solution of those problems involved a technique that was new to the students. The question, then, was whether to equalize the time spent on problems (the result most likely being that the software groups would be able to complete more problems) or to equalize the number of problems completed (the result being that the software groups would spend less time on the problems). When the time either way, however, was limited to one hour it was decided that the full hour should be utilized for all sections.

With that in mind, a set of problems was developed for the students of each section to work on. The set was sufficiently long enough that few students would complete it with confidence. The problems fit on one sheet of paper, front and back, and asked nine questions about two different trusses, see figure 6.3. The trusses were chosen for specific reasons. The Howe truss (the triangular shaped one) was chosen because it is the truss used on the simulation page for the method of joints section of the content-type software, Multimedia Engineering Statics (see Chapter 4). This, however, did not necessarily create a bias towards the content-type group because the truss could also be created easily in the tool-type software and is one that is often solved by hand when learning truss analysis. The cantilevered truss was one that some instructors specifically mentioned as being illustrative in that it is an excellent lead-in to the analysis of beams. This truss could be constructed in both the tool-type and content-type software environments and could be solved with relative ease by hand and as such did not lend itself to one particular method either.

For the following exercises, consider the truss in the figure below, which is support by a pin at H and a roller at D.



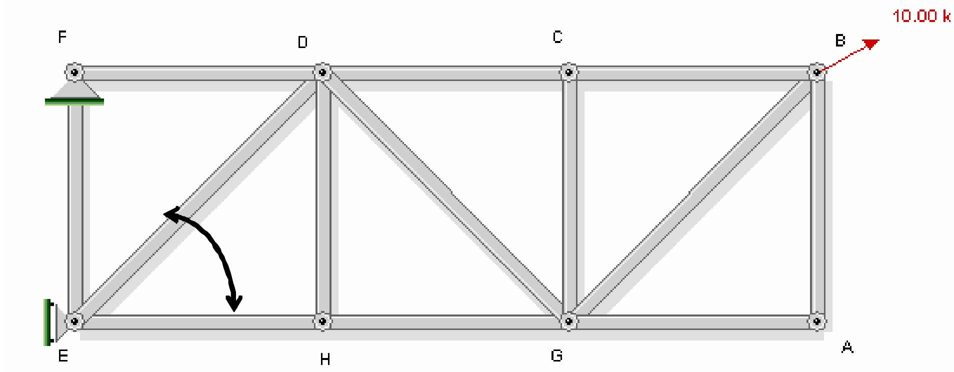
If ? = 45° and a 9kip force is applied vertically at *A* in the downward direction, determine the forces in GH, CH, and AF and indicate whether the member is in tension or compression.

What is the relationship between ? and the force in members CH and GH?

Do any of the member forces change if the 9kip force is applied at *F* rather than *A*; if so, how?

How do the member forces change if the 9kip load is distributed across the bottom cord: 3kip loading at each of the joints *G*, *F*, and *E*?

Figure 6.3a Special Session Problems



Consider the cantilever truss in the figure above which is supported by a pin at *F* and a roller at *E* and has an applied load of 10kips at *B*; if ? is equal to 30°, identify which members are in tension, which members are in compression, and which are zero force members.

Which members have the maximum compressive and tensile forces?

If ? is 0°, which members are in tension, in compression, or zero force members.

If ? is 270°, which members are in tension, in compression, or zero force members.

If ? remains 270° but member BG is removed and replaced by a new member AC, how does this affect the force in AB? Does this change in configuration affect the maximum compressive and tensile forces in the truss?

Figure 6.3b Special Session Problems (cont.)

The questions included some quantitative analysis problems and some qualitative truss analysis questions that statics instructors identified as important (see Chapter 4). Some of the problems asked students to determine how changing certain parameters of the truss would change the behavior of the truss as a whole. As mentioned earlier, this is precisely what instructors hoped that software use would be able to help students with; working with the truss as a complete structure rather than the analysis of a series individual elements. The posttest as described in Chapter 5 tested students abilities to both quantitatively and qualitatively analyze trusses and so the special session focused on those two abilities as well. The results of the posttests, described in the following chapter, will break down the scores by type of question to determine if the groups performed differently on types of questions (qualitative vs. quantitative) as well as if there was a difference in overall scores. The statics instructors previewed the instrument and approved it for use in their courses.

The content-type group met in a CEE computer lab for the special session. They were directed toward the MES website and instructed to thoroughly investigate the site including all the movies, audio clips, and simulations. They were given the problem set and told to complete as many of the problems as they could in the allotted time period. They were allowed to work in teams and given headphones for the purpose of listening to the clips and movies. An interesting thing happened about ten minutes into the special session. Once one student realized that they could use the simulation pages to complete the assignments, many of them skipped through the content pages quickly in order to get to the simulation pages and therefore complete this assignment. This despite the fact that they were informed that the assignment was neither going to be graded nor count towards their five percent participation grade. All students did, however, spend some time on the content pages although the amount of time spent on those pages varied greatly. There was one problem on the assignment that students had to complete by hand, observations of students' responses revealed that most of the students guessed on this problem because there was no scratch work or computations and most of the answers were wrong. Most of the students completed the assignment though they guessed on the aforementioned problem rather than spending time to work it out.

The tool-type group also met in the CEE computer lab. They were given a brief introduction by the instructor about how to start the truss environment and then how to build, constrain, and load a truss. They were then given the problem set and asked to complete the problems through the use of the content-type software, Dr. Frame. In Dr. Frame the students were allowed to build each of the two trusses and then adjust some of the parameters to see how these adjustments affected the truss, as with the simulation pages on MES but in a more powerful and easier to use environment. Despite being easier to use than other structural analysis programs, Dr. Frame did have some glitches that were not revealed until over forty students were using the program all at once. Consequently, the researcher spent much of the special session answering questions about how use the software, which was an anticipated occurrence and did not interfere with the session as planned. Again, students were allowed to work in pairs or teams. Some of the students were able to complete the assignment but others were not. The reason for this is that, despite being easy to use, the program did take a few minutes to learn and become accustomed to which did take some time away from completion of the assignment.

The comparison or control group met in their regular classroom. They were given the problem set and asked to complete as much as possible in the hour that had been given them. They were also encouraged to work in groups as the researcher walked around the classroom answering questions about truss analysis in general and the problems specifically. Very few of the students completed the assignment, and only a few even completed one whole page. Clearly working the problems by hand was a disadvantage in terms of time during the special session. Per the decision mentioned above, however, the concern was not to have the students complete the same number of problems; the concern was to have the students work on problems for the same amount of time. And this situation is actually authentic. In a real-world classroom, there is a limited amount of time available both in and out of class in which to practice skills or complete assignments and if using an instructional software allows students to work through more problems and gain more experience in the same amount of time as it would take to complete fewer problems perhaps this is a real benefit. As such, it was not seen as a bias or an unfortunate consequence, but an authentic, real-world, and valid occurrence. Also, as the posttests and exams would be completed by hand, as they are in most real-world situations, perhaps completing a few problems by hand would better prepare the comparison group for future assessments than would completing many problems through the use of a computer. The results of those future assessments are presented in the next chapter and they will reveal if there was a difference and may lend some insight as to why.

The problem sets for each of three groups were collected but they were not graded. There are a few reasons for this. First, because the students were encouraged to work in pairs, some teams only turned in one set of completed problems for both students rather than one for each of the individuals. This was acceptable because the purpose of the special session was not to have them complete the assignment; the assignment served as a guide to help them apply their skills within set parameters. Completing only one assignment per group allowed the students, ideally, to focus not on writing down the solutions but how the solutions were obtained. Second, the problems completed during the special session were not intended to be assessment instruments in the original research design and every attempt was in this project to stick to the original design. Third, it was assumed from the beginning that the comparison group would not be able to complete as many problems as either of the software groups and as such comparing the groups on these problems would not be illustrative. The problems were collected simply for review by the researcher for general observations. The only major observations that resulted from this review were how many of the problems the different groups were able to accomplish and the aforementioned situation where a majority of the content-type students did not spend time to work out a problem that required a hand solution.

The special sessions seemed to have been conducted successfully. There were little concerns aside from a few computers not working properly and the session being shorter than originally proposed. The results of the analysis of the assessment data as presented in the following chapter reveal whether or not the special session had an actual effect on the learning process.

6.2.1.5 Posttest

The posttest was completed in the lecture period following the special session. The design of the posttests is described in an earlier chapter of this document and the final form of the posttest is shown in Appendix C. A cover page was drafted for this instrument as well, which can also be seen in the appendix, though much shorter in form. The cover page was shorter because students did not need as much motivation to complete the assignment because they were going to be completing the posttest in class and thus needed no extra prodding to complete the assignment; the assignment was the class activity that day. Also, the researcher, who disseminated the posttest, gave all the verbal instructions that were necessary to complete the exam. Finally, timeliness was important as the posttest was being completed in class and so the cover page was shortened so that students could spend more time completing the posttest and less time reading the cover page.

The students in each of the sections were allowed approximately 40 minutes to complete the posttest, which for many students was not enough time to complete the assignment. Differential completion rates, however, were intended in the design of the posttest and this effect will further illustrate differences in abilities between the different subject groups. While the students were completing the posttest, the researcher wandered the room and answered what few questions students had about the instrument. When the 40 minutes had passed, the researcher collected the posttests, the grading and analysis of which will be described in the following chapter.

6.2.1.6 Exam Questions

Once the intervention, and therefore the topic of trusses, was completed it was necessary that none of the instructors addressed truss analysis at any point during the remainder of the semester so as not to bias retention assessments that would take place in the future. Both instructors understood the importance of this matter and agreed not to address trusses at all throughout the remainder of the course with the exception noted below. Both Instructors A and B had the same homework policy; they both assigned homework problems but not collect or grade the problems. They did give students the answers to the homework problems so they could check their work, but they were never held accountable for their homework. With this in mind, the same homework problems were assigned to all sections involved in the project and the students were given the answers but the homework was never collected or graded.

While not addressing trusses in class, the instructors did help students outside of class in the form of help sessions and one-on-one assistance during offices hours. Some students also went to sources outside of class for private tutoring. While there is no reason to assume that the number of students who sought outside help for trusses, students were nevertheless asked a few questions on the retention test (discussed in the next section) about how often and from where did students seek outside help on the topic of trusses. The results of these questions concerning outside help were compared between the three groups and are presented in the following chapter.

The one time that instructors did address trusses again in class was on a midterm exam, where truss analysis abilities are normally assessed in CEE 2020. In order to avoid any bias on the future retention tests, Instructors A and B agreed to put the same truss analysis question on each of their midterm exams so that each of the students would be exposed to the same problem. Due to a miscommunication however, slightly different analysis problems were used by the different instructors. The two problems differed slightly in the truss configuration, one had more members than the other which also affected the reaction forces, but the solution methods were identical.

The two problems, see figures 6.4 and 6.5, though only slightly different had the potential to create a bias between the groups if the differences were significant enough. To alleviate this concern, the researcher gave both exam questions to the students in his own class (Section F), which were not a part of the experiment, to determine if there was a significant difference between the two instruments. Students completed both truss problems in an exam setting like the midterms taken by the research groups and the problems were graded and scored by the researcher. Great care was taken during the grading process to ensure that each of the problems were graded according to the same rubric and by the same standards.

The truss shown below is supported by a pin at *A* and a roller at *D* and has a load of 20kN applied at *F*. Find the reactions at *A* and *D*, then use method of joints to find the force in member *AB* and method of sections to find the force in member *EF*.

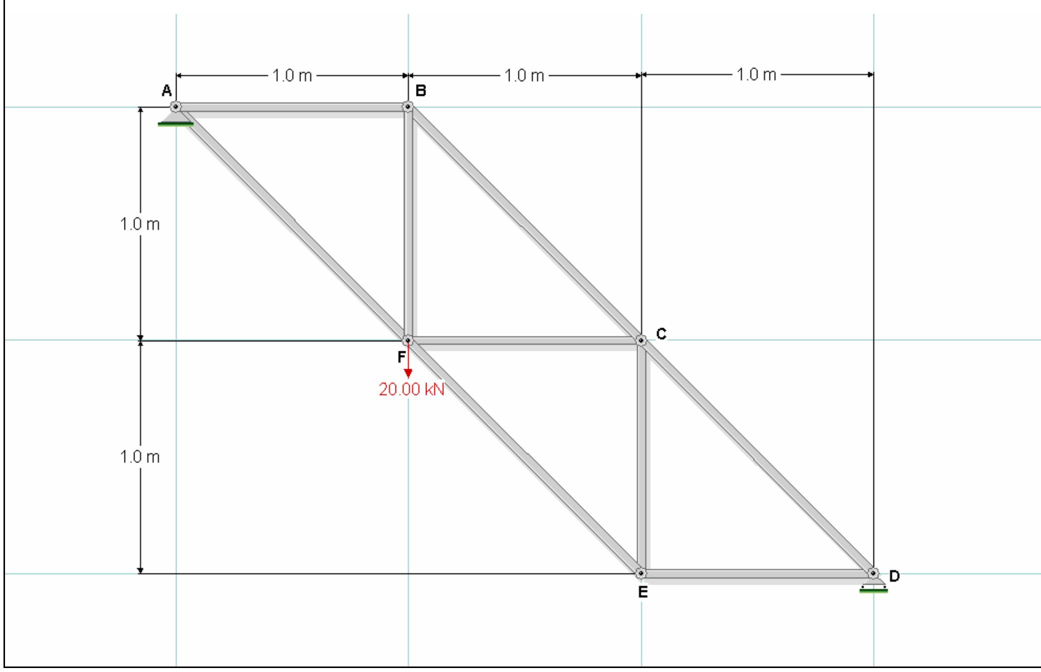


Figure 6.4 Truss Question Included on Midterm Exam in Section H of CEE 2020, Fall 2002.

The truss shown below is supported by a pin at *A* and a roller at *E* and has a load of 20k applied at *H*. Find the forces in members *AB* and *HG* and state whether they are in tension or compression.

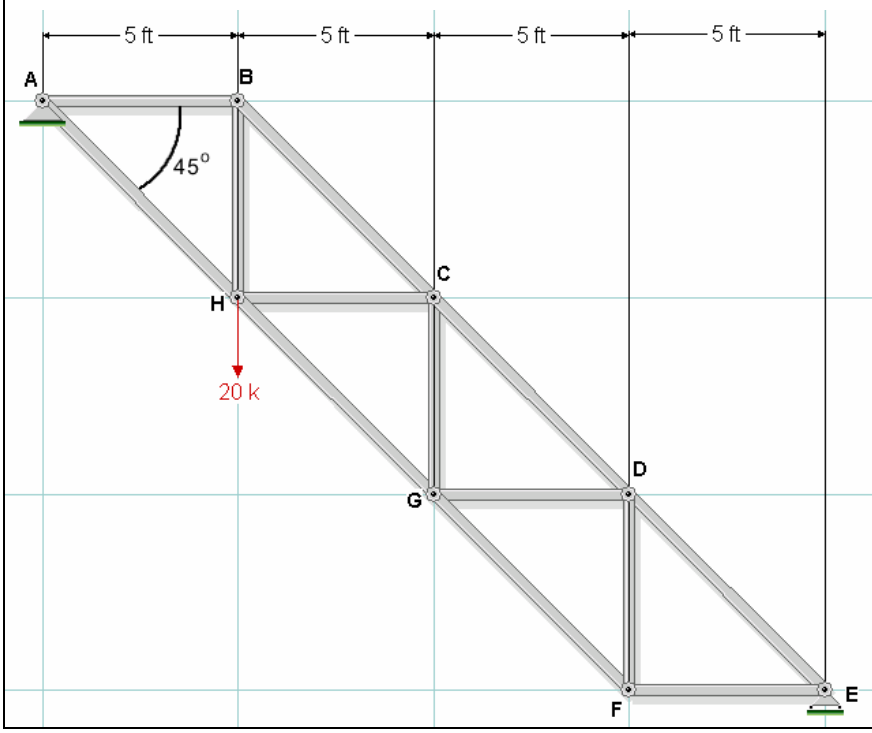


Figure 6.5 Truss Question Included on Midterm Exam in Sections C and E of CEE 2020, Fall 2002

A reliability analysis was performed on the results from the two problems taken by the students in Section F. Both problems were scored out of a total of 25 possible points. The mean for the first problem was 20.12 and the standard deviation was 4.88. On the second problem, the mean was 20.29 and the standard deviation was 5.45. Three different analyses were performed, the first of which was a correlation to see if the results from the two forms were related. This analysis yielded a correlation coefficient of r=.6115, which was significant at the .01 level. This result indicates a significant, positive relationship between the two forms, which is to say that there is a significant relationship between a student's score on one problem and his or her score on the other problem. Further, a split half reliability analysis yielded a correlation coefficient of r=.7560 between the two forms. This result also indicates significant relationship between the two forms; it indicates that the two forms are measuring the same thing and may in fact be two halves of the same form. Finally a Wilcoxon signed rank test was used to compare the means from the two problems, similar to a t-test but for distributions that may not be normal, and revealed that there was no significant difference between the means (Z=-0.068, p=.946) on the two exam questions. All of this lends evidence to the fact that despite the differences between these two exam questions, they essential test the same thing, in the same manner, and to the same extent. As such, the exam questions can considered to be the same.

With evidence to support this consideration, there was no longer any concern about the two different types of exam questions. Furthermore, the exam questions were used as another assessment in the study. Because the exam questions were not significantly different, the results could be compared between the groups to see if significant differences did occur between groups due their experimental conditions. All sections completed the exam questions approximately one week after the posttest and the instructors made copies of the students exam questions and gave them to the researcher for use in this project. The exam questions were scored by the researcher and the scores were analyzed for differences between groups. The results of this analysis will be discussed in the following chapter.

6.2.1.7 Retention Test

The intervention concluded the third week in September, 2002. Approximately ten weeks later, during the last week in November, the retention test was administered to the three research groups. This test was administered during "dead week" at Georgia Tech, which is the last week of classes before final exams; a week in which no exams are to take place. Naturally, this is very busy time for instructors who are trying to wrap up their courses and prepare students for final exams. As such, it was not feasible to have students complete the retention test during class and so it was given as a take home assignment.

Once again, a cover page, shorter than the one for the pretest and shown in Appendix E, was attached to the assignment informing students as to the importance of their participation and honest effort on the assignment. The assignment itself was identical in form to the posttest and students were asked to take no more than forty minutes, the amount of time given to work on the posttest, to work on the assignment and that after such time they should stop working even if they were not done. In this way, every effort was made to assure that the retention test was as close to the retention test as possible so that differential scores could be measured to assess what students remembered after ten weeks.

As with the pretest, students were given five days to complete and return the assignment. Along with the retention-test, students were asked to answer some basic demographic questions as well as questions about their experiences using the software if they were in one of the groups that used the software (see the questionnaire with the cover letter in Appendix E). Also on this demographic questionnaire were the aforementioned questions about whether or not students sought help outside of class with trusses or other topics and if so, where did they go for the help. Students were informed that completion of this assignment did count towards the five percentage points that they would receive on their final course grades for participating the in research project.

6.2.1.8 Long-term Retention Test

The long term retention test was administered in March of 2003, approximately fifteen weeks after the retention test. By this time, some students were eleven weeks into the following semester and others had graduated. As mentioned in the first half of this chapter, the decision was made to administer the long-term retention test to CEE students only and that the best way to contact and motivate students to participate would be to hand the assignment out in a subsequent course.

Two courses, CEE 3030, Strength of Materials, and CEE 3020, Civil Engineering Materials, are required to be taken simultaneously and are normally taken in the semester following CEE 2020. The instructor of CEE 3020 agreed to help with the project by allowing the long-term retention test to be handed out in that class. There was insufficient time to complete the assignment in class (especially for a topic that was not directly related to the course), but the instructor allowed the assignment to be given as a take-home. Further, the instructor agreed to provide motivation for the student to complete the assignment by again giving them five points towards their final grade in CEE 3020 if they returned the assignment by the due date.

Not all of the CEE students who participated in the first phase of the project were enrolled in CEE 3020. Those who were not, received the long-term retention test via email and asked to participate and return their completed tests via email or turn them into the researchers CEE mailbox. Very few people who were not in CEE 3020 chose to participate even though they were emailed a number of times and politely asked to contribute. Exact participation numbers for students both enrolled in CEE 3020 and otherwise are given in the following chapter.

The long-term retention test was identical in form to the posttest and retention test. Again, students were asked to work for no more than forty minutes on the assignment and to not use any books, notes, or other resource besides a calculator. A lengthy cover page, similar to the one attached to the pretest and shown in Appendix F, explained anew the importance of the project and the participation of the students. The collection of the long-term retention tests completed the initial implementation of the intervention. Even as this implementation was concluding, however, the replication study had already begun.

6.2.2 Replication, Spring 2003

In order to preserve the validity of this quasi-experimental research project, the intervention was conduced again, or replicated, in the spring of 2003. For the most part, the replicated study was identical to the initial study; a few changes were made and they will be discussed below.

6.2.2.1 Subject Selection

In the spring of 2003, seven sections of CEE 2020 were taught by five different instructors, two of which taught two sections of the course. As with the initial study, upon the close of late registration data were gathered on each of the sections to determine which sections of the course had the highest number of CE students enrolled in them.

These data are presented in table 6.4, which shows the number of CE students and the number of total students for each section of the course taught in the fall of 2002. Sections B, E, and G were chosen to be the three different sample groups for the study because they contained the greatest number of CE students.

Table 6.4 Number of CEE Students in Each Section of CEE 2020, Fall 2002.

Section	CEE Students	Total Students
B	16	29
C	2	45
D	6	26
E	21	46
F	1	14
G	9	48
I	2	40

Sections E and G were taught by Instructor A who agree to once again let the intervention take place in those sections. Instructor B did not teach CEE 2020 in the spring of 2002; section B was taught by another instructor, referred to here as Instructor C, who agreed to let the intervention take place in that section as well. Again, the reasons for these sections having the higher CEE enrollment rates were not coincidental. Instructor A once again limited enrollment to sophomores only during early registration. A similar situation occurred in section B, which was limited during early registration to CEE students only, though it can be seen that during late registration many other majors signed up as well.

Section B met on Mondays, Wednesdays and Fridays from 9:00 A.M. to 10:00 A.M. and was assigned to the Content subject group, or the group that would use the content-type software during the special session of the intervention. Section E met on Mondays, Wednesdays, and Fridays from 12:00 P.M. to 1:00 P.M. and was assigned to the Tool subject group, or the group that would use the tool-type software during the special session of the intervention. Section G met on Mondays, Wednesdays, and Fridays from 2:00 P.M. to 3:00 P.M. and was assigned to the comparison or control group, the group that would solve problems during the special session without the assistance of any software. As before, these assignments were made for no particular reason, yet they were not randomly assigned either, see section 6.2.1.1 for more explanation on this matter.

More detailed information about the sample groups will be presented in the results chapter, which will describe various demographic data about the groups and make comparisons between the groups based on these demographics.

6.2.2.2 Pretest

The format of the pretest did not change, nor did the manner in which is was administered change. It was, once again, a take home assignment. It was handed out on the first day of the intervention and collected five days later. Once again, there were very few questions or concerns from the students regarding the pretest.

6.2.2.3 Instruction

Little change took place in the instruction portion of the intervention either. The only difference is that all three sections were Monday-Wednesday-Friday classes whereas in the initial study, one section was a Tuesday-Thursday class. There is no reason to assume that this change would have any effect on the study or its results. The lectures from the previous semester were saved and given again to the students in the replication study, so that students in all six sections received the same instruction. Also as before, the researcher taught all three sections during the intervention.

6.2.2.4 Special Session

The special session during the spring happened exactly as it had during the preceding fall. The content-type group, section B, met once again in the computer lab and completed the problem sets in figure 6.3 with the assistance of MES. The tool-type group, section E, also met in the computer lab and completed the problems with the assistance of Dr. Frame. The comparison group, section G, met in class and worked on the problems by hand. All groups were again encouraged to work in pairs.

One observed difference was that the content-type group spent more time on the content pages and did not rush to the simulation pages as the students had done in the previous semester. The content-type group was still hesitant to complete the one problem that needed to be done by hand and many of the students simply guessed as their fall counterparts had done. Also as before, the comparison group was not able to finish all the problems by hand.

6.2.2.5 Posttest

The posttest used in the replicated study was identical to the posttest used in the initial implementation and it was administered in an identical manner as well. Students were given the posttest in the lecture period that took place after the special session. They were allowed forty minutes to complete the assignment in class after which the class was turned back over to the instructor who moved on to a new topic.

6.2.2.6 Exam Questions

Once again, instructors agreed to not address the topic of trusses in class for the remainder of the semester. Instructor A followed the same homework format as before (assigning it but not collecting or grading it) and Instructor C did as well for the truss topic; though students in this section were normally held accountable for their homework, Instructor C changed the homework format during the intervention to match Instructor A's format. The same homework problems were assigned to each of the sections and they were all given the answers to the homework problems, which were not collected. Some students again sought help outside of class and the extent to which this was done, as well as where they sought help, was assessed on the retention test.

Similar to the initial study, truss questions were included on a midterm exam. This semester, however, Instructor A wished to have two truss questions on the midterm. Also, Instructor A did not wish to use the same questions as had been used the previous semester because students from the previous semester had solutions to those problems. Finally, Instructor A wanted to have significant input on the design of the exam questions this semester so they were somewhat different than the exam questions from the previous semester. As such, two new exam questions were developed and administered on a midterm exam to all three sections involved in the study. These two questions, shown in figures 6.6 and 6.7, were quantitative analysis questions, one of which was to designed to be solved using method of joints while the other was designed to be solved using method of sections. Because the exams were so different from semester to semester, an because no reliability analysis was done between new and old exams, no comparison could be made across semesters with regards to the exam questions. They were, however, still a valuable tool for comparing different groups within each semester.

1) The truss shown below is supported by a pin at *O* and a roller at *A* and is loaded as shown at *I*. Determine the forces in members *ED*, *DK*, and *IJ* and state whether the members are in tension or compression.

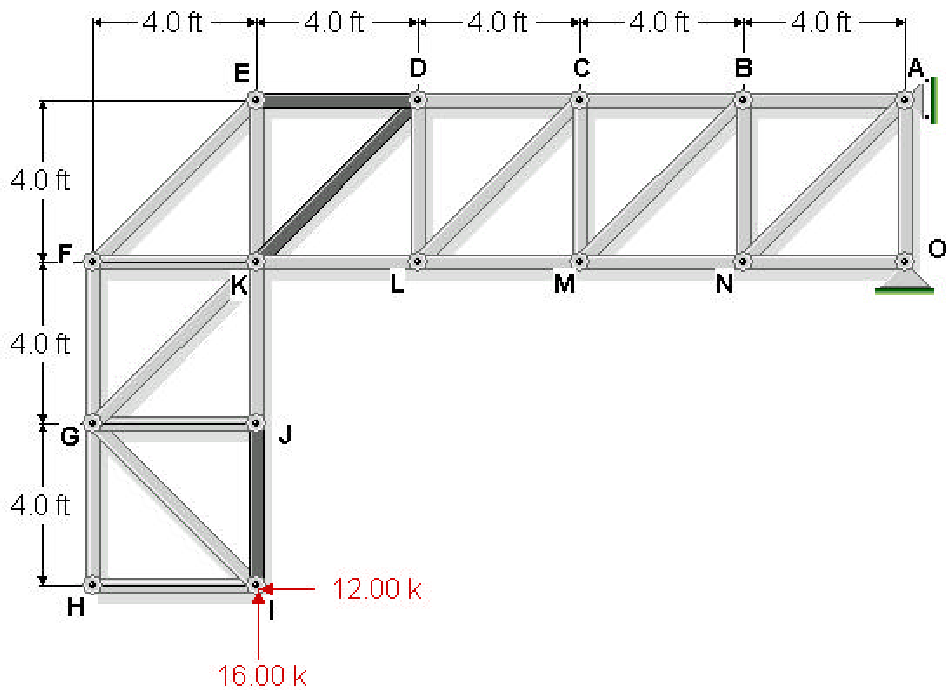


Figure 6.6 Exam Question from Midterm Administered to All Research Sections, Spring 2003

1) For the inverted Fink Truss below, which is supported by a pin at A and a roller at E, find the forces in members HB and FD and state whether they are in tension or compression.

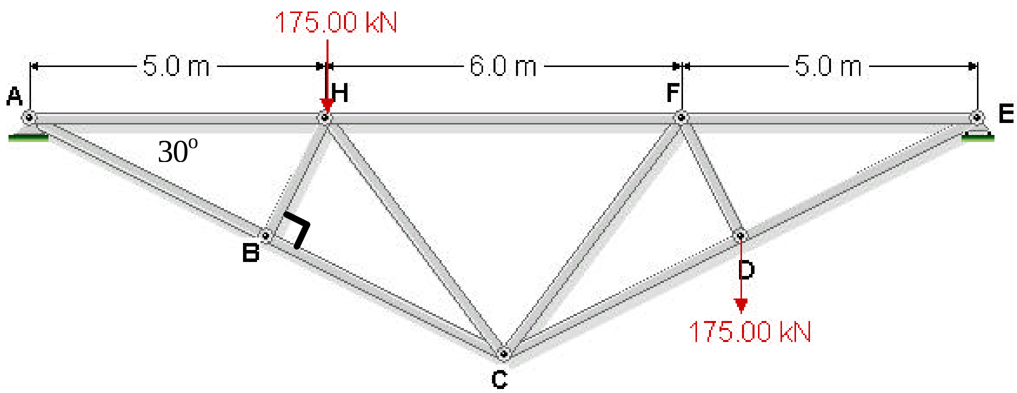


Figure 6.7 Exam Question from Midterm Administered to All Research Sections, Spring 2003

6.2.2.7 Retention Test

The intervention was completed in February and the retention test was administered approximately ten weeks later in late April, once again during dead week at Georgia Tech. The retention test administered in the spring was identical in form to the one administered in the fall. Students took the test home and had five days to complete it. They were once again asked to work on the assignment for no more than forty minutes at which time they were to stop.

6.2.2.8 Long-term Retention Test

The long-term retention test was identical in form to the one administered in the previous implementation and it was administered in the exact same manner, approximately twenty-five weeks after the intervention. The instructor for CEE 3020 agreed to hand the test out in class as a take home assignment and give students points toward their final grades for returning the completed test. Once again, all students who participated in the intervention but were not in CEE 3020 were emailed a copy of the test and asked to complete and return it. As before, the return rates for the emailed exam were very poor, exact numbers will be given in the results section.

6.3 Summary

This chapter began with a list of changes that were made to the original proposal based on input from the instructors and students, the second half of this chapter explained exactly how the intervention took place in both of its phases. For the most part, the intervention was carried out exactly as proposed and with very few glitches. Data were collected on prior knowledge, learning, retention, and long-term retention. Unexpectedly, midterm exam questions were also collected and provided another assessment of learning. With all this data collected and the interventions complete, the time came for the next phase of the project: the data analysis. The analyses of the data and the results of these analyses are the topic of the next chapter.

CHAPTER 7

RESULTS

To determine whether the intervention had an effect on performance or retention, an analysis of the data gathered from the various assessment instruments was completed. In performing such an analysis, it is not sufficient to simply compare scores side by side and declare that the section with the highest scores was made up of students who performed better than those in the other sections. The difference between scores could simply be the result of random effects or chance.

In order to ensure that an actual effect occurred as a result of the independent variable (or intervention) two things must be accomplished. First, an appropriate research design must by chosen or designed and then followed as precisely as possible in order to rule out any undesirable, non-experimental variables that may unintentionally affect the results. Second, a rigorous and methodical statistical analysis of the experimental data must be followed in order to assure that the effects are not merely random in nature. This document, up to this point, has been focused mainly on the former requirement: that a proper methodology has been adhered to so as to rule out any non-experimental effects. The focus now shifts and becomes statistical in nature, to rule out the second type of undesirable effects: those that are random.

This chapter will discuss the analysis of the data from this experiment and will present the quantitative results. A discussion of the findings based on these results, however, will be reserved for the next chapter. More specifically, this chapter will begin by restating the goals and hypotheses of this research to give direction to the analysis process. Next, a brief review of statistics in general, and the methods used in this research specifically, will be presented for the benefit of those not familiar with the analysis methods that are commonly used in the behavioral sciences. Finally, the analyses and results will be presented in order of the instrument used and the semester in which the data were collected.

7.1 Review of Research Objectives

As the analysis of the data will be used to support or reject the hypotheses as stated in chapter 3, it is beneficial to restate those hypotheses here in order to understand the purpose of the analysis methods. The hypotheses have been revised slightly to reflect revisions in the methodology and to included terminology that has been subsequently introduced. The hypotheses are also stated in a manner that relates quantitatively to a specific assessment instrument, which is appropriate for a data analysis section. For the rationale behind these hypotheses, please refer to section 3.3. The hypotheses are as follows:

1. The experimental groups will perform significantly better on the posttest than the comparison group.
2. The experimental groups will perform significantly better on the exam questions than the comparison group.
3. The experimental groups will perform significantly better on the retention posttest than the comparison group.
4. CEE students from the experimental groups will perform significantly better on the long-term retention test than CEE students from the comparison group.

5. There will be no significant difference in performance on the posttest between the content-type and the tool-type groups.
6. There will be no significant difference in performance on the exam questions between the content-type and the tool-type groups.
7. There will be no significant difference in performance on the retention posttest between the content-type and the tool-type groups.
8. There will be no significant difference in performance on the long-term retention posttest between CEE students in the content-type group and CEE students in the tool-type group.

Of course, more information will be sought from the data to answer other questions as well. For instance, pretest data will be used to determine whether the groups differed in terms of prior knowledge. In addition, results may be broken down based on major, rank, or other demographics. All this will be described in detail in the following sections as will the results of the analysis. These eight hypotheses, however, were the main objectives of the research and all other results will either support these in some way or will be ancillary to them.

7.2 Review of Statistics

Because the statistical analysis of the data will form the basis upon which inferences will be drawn concerning this research, a very brief discussion on statistics common to the behavioral or social sciences is given here, mainly for the benefit of some interested engineers who may not be familiar with all the methods used herein. The theory of statistics is divided into two major parts or functions: descriptive statistics and inferential statistics (Hays, 1994).

Descriptive statistics are used to effectively describe or summarize large amounts of data (Hays, 1994). One common example of a descriptive statistic is a students grade point average (GPA), which summarizes a student's performance in courses throughout school into a single number. While looking at a student's GPA may not be as illustrative as looking at a complete transcript, it is an efficient way of gauging that student's abilities.

Inferential statistics, on the other hand, illustrate how data can be used to infer or draw conclusions about a large population of possible measurements based upon data drawn from a small sample of that population (Hays, 1994). Inferential statistics allows for generalizations about larger groups, such as all engineering students in this case, to be made from measurements drawn from far fewer subjects or measurements, students in six sections of CEE 2020 in this project (Sirkin, 1999). As the ultimate purpose of this experiment is not to merely describe what happened in the six subject groups involved in the experiment, but to generalize their experiences to the larger population of engineering students who use software in education, inferential statistics were used.

Furthermore, inferential statistics can be broken down into two different types: parametric statistics and nonparametric statistics. Parametric statistics are based upon certain assumptions about the data, most notably that the data are drawn from a normally distributed population (Siegel and Castellan, 1988). When such assumptions about the population data are unfounded, nonparametric techniques should be used. Nonparametric tests are more generalizable because they are not based upon rigid assumptions that may not be true and are usually not tested (Siegel and Castellan, 1988). Both parametric tests and nonparametric tests were used in this project. Listed below, following some general terms and definitions, are the statistical tests that were used in this project categorized as being descriptive, parametric or nonparametric. Brief descriptions of each of these statistical tests are also given.

<u>7.2.3 General Terms and Definitions</u>

There are a few general terms and definitions that will be used throughout the remainder of this chapter and are defined here for clarity (source: Hays, 1998).

- **Null Hypothesis**: The quantitative hypothesis that is actually tested in the statistical test. The null hypothesis may not always be the same as the substantive hypothesis. For example, to test for a difference in mean performance between two groups, a t-test may be used where the null hypotheses is that there is no difference between the two means. Rejecting the null reveals that there is a significant difference.
- **Significance level or p-value (p)**: The probability that a statistic would be as extreme as the one observed (i.e. calculated) if the null hypothesis were true. This is often loosely interpreted to mean that p is the probability of the null hypothesis being true. Though this interpretation is inaccurate, it is illustrative. When p is low and approaches a pre-specified significance level, or alpha level (a), the null hypothesis is rejected. The alpha level is usually .05, as in this project, or .01.
- **Types of Data**: Four general types of data or measurement levels exist and the statistical test that can be used depends on the type of data measured. They are as follows:
    - **Nominal**: Non-quantitative, categorical data such as gender.

- **Ordinal**: Non-quantitative, categorical data to which there is a logical order of categories. Likert rankings (strongly agree, agree, neutral, etc.) are common examples of ordinal data.
- **Interval**: Continuous, quantitative data where the intervals between adjacent measures are equal. Furthermore, interval measures have no true zero. A classic example of interval data is temperature measured in degrees Fahrenheit. The difference between 50° and 60° is the same as the difference between 90° and 100° but it is incorrect to say that 36° weather is twice as hot as 18° weather.
- **Ratio**: Continuous, quantitative data similar to interval but with a true zero. Weight measured in pounds is an example of ratio data. A 30 lb stone weighs three times as much as a 10 lb stone.

7.2.4 Descriptives

Descriptive statistics are familiar to most people and are often encountered in daily life. Descriptives also form the basis for many inferential statistics (Hays, 1994). For the sake of completeness, the descriptive statistics used in the analyses that follows are listed and described here (source: SPSS Inc, 1998).

- **Mean**: The mean is a measure of central tendency. It is the arithmetic average, the sum of a set of numbers divided by the number of numbers in the set.
- **Median**: The median is another measure of central tendency. It is the number that falls in the middle of an ordered set, that is to say that half of the numbers in the set are larger than the median and half the numbers are smaller.

- **Variance**: Variance is a measure of dispersion or how far from the mean the data are spread. It is calculated by squaring the difference between each number in the set and mean and then averaging these squares.
- **Standard Deviation**: Another measure of dispersion, the standard deviation is the square root of the variance. In a normal distribution, 68% of the numbers in the set fall within one standard deviation of the mean and 95% fall within two standard deviations of the mean.
- **Range**: The range is the difference between the minimum, the smallest number in the set, and the maximum, the largest number in the set, and is another measure of dispersion.
- **Confidence Interval**: A range of values with a given probability of covering the true population value (Hays, 1998). For example, a particular measure may yield a *sample* mean of 44 with a 95% confidence interval of 37-48. This suggests that there is a 95% chance of the *population* mean actually falling between these two values.

7.2.5 Parametric Statistics

Most parametric statistics are based on the mean and standard deviation (i.e. the parameters). These parameters are not appropriate for all types of data. Even if the data type is appropriate, it is important to note that parametric statistics also require certain assumptions, usually requiring that the distribution of the data be normal. If there are good reasons to assume normality or if tests are done to reinforce the assumption of normality, parametric statistics are appropriate for use in hypothesis testing. Some tests, such as the Analysis of Variance (ANOVA), are robust even if data are not normal but may still be based on other assumptions. Listed below are the parametric statistics that were used in this project accompanied with descriptions, assumptions, and requirements of each (source: SPSS Inc, 1998). As a side note, most inferential statistics require that the observations or measurements are independent and take from a random sample. As this research followed a quasi-experimental design, randomness was not assured and thus a selection bias could have been introduced. As described earlier, however, this selection bias is overcome in this project through replication and so this assumption will not be mentioned again.

- **Student's t test**: The t test is used to either compare two means or to compare a mean to a known value. This test assumes that the data are normally distributed but is robust to normality given that the distribution is at least symmetric. The data must be quantitative (interval or ratio). The null hypothesis is that the means, or the mean and the known value, are the same. The t statistic is calculated and compared the known probability distribution of t and the null is rejected if $p<a$.
- **Analysis of Variance (ANOVA)**: ANOVA is similar to the t test except that it can compare multiple means as opposed to just two. As with the t test, ANOVA assumes that that data are quantitative and normally distributed, though the test is fairly robust as long as the distribution is symmetric. Furthermore, the test assumes that the variances associated with the means that are being compared are equal. The null hypothesis is that all means are equal. The F statistic is calculated and compared to a known probability distribution of F and the null is rejected if $p<a$. If more than two groups are being compared, ANOVA will simply indicate that a difference exists, it does not identify where the difference

occurs (e.g. 1, 2, & 3 are not all the same but is 1 different from 2, or 2 different from 3, or 3 different from 1). Further tests, called post-hoc comparisons, are required to identify the nature of the difference.

- **Bonferroni Comparisons**: This post-hoc comparison technique uses the t test to compare pairs of means. In this technique, an adjustment is made to the significance level based on the number of pairs compared. To illustrate this technique, imagine an ANOVA was completed with five different conditions and the null hypothesis was rejected, meaning that at least one condition was significantly different than the others. Ten t tests would need to be completed to determine where differences occurred (compare: case 1 to case 2, case 1 to case 3, case 1 to case 4, case 1 to case 5, case 2 to case 3, etc.). If each of these comparisons is done at the .05 significance level then there is a one-in-twenty chance of an individual test being incorrect. If ten tests are done one the same data at this significance level, then there is a forty percent chance of one test giving incorrect results. To alleviate this, the Bonferroni method divides the desired, or familywise, error rate (a_{FW}) by the number of comparisons to get a pairwise significance level (a_{PW}) which is then used for each of the t tests. In this example, if $a_{FW} = .05$ and 10 comparisons were made, $a_{PW} = .005$, which is the significance level that would be used for each of the ten t test comparisons.
- **Tukey Honestly Significant Difference (HSD)**: The Tukey HSD method is another means of conducting post-hoc comparisons, which reveals significant differences between means similar to the t test. The Tukey method corrects for the familywise error rate while simultaneously making the comparisons as

opposed to the Bonferroni method which adjusts the error rate and then uses t test in a two step method. The Tukey HSD is more powerful when many comparisons are being made but the Bonferroni method is more powerful when just a few comparisons are made. Other post-hoc techniques are available, but these are the two most widely used methods and they were the ones chosen for this project.

- **Pearson's Correlation Coefficient (r)**: A measure of the linear relationship between two variables, r ranges from -1 to 1. The greater the absolute value of r, the stronger the relationship. A positive value indicates a direct relationship and a negative value indicates an inverse relationship (i.e. as one variable increases, the other decreases). This type of correlation requires symmetric, quantitative variables.
- **Planned Comparisons:** In some cases, instead of using an ANOVA to compare all means, planned comparisons may be used to test specific, predetermined hypotheses about the data. The advantage of using planned comparisons is that if they are designed correctly, the pairwise comparison may be used as opposed to the familywise error rate. Unfortunately, the number of independent planned comparisons that can be made with k samples of data are k-1, meaning that planned comparisons may not be used to compare all samples. If comparisons between all samples are desired, it is more appropriate and accurate to use an omnibus test (such as ANOVA) followed by post-hoc tests. Planned comparisons are implemented by multiplying means by pre-specified coefficients and summing them into two groups which are then compared using a t test.

- **Difference of Proportions Test:** This test is used to compare two percentages or proportions to determine significant differences. This test is quite similar in form to the t test except that instead of comparing two means, two percentages are compared. In this test, a Z-score, or standardized normal test statistic, is calculated and compared to the known standard normal distribution. The null hypothesis is that the proportions are the same. This test assumes that the proportions have a normal sampling distribution.

7.2.5 Nonparametric Statistics

Nonparametric statistics are used in a number of different situations. One previously mentioned case is when the assumptions intrinsic to the parametric tests are not met. Another reason they are often used is if the data are not continuous or quantitative. In such cases the parameters, the mean and standard deviation, upon which parametric tests are based, are no longer appropriate. For these reasons, some nonparametric tests are based on the median. Some nonparametric techniques are also ranking tests or ordered tests, which focus not on the numerical values but on the ranking of the scores; these tests are useful for ordinal data (Siegel and Castellan, 1988). Still other tests are available for nominal data that do not rely on any type of order. Below are listed and described the nonparametric statistical techniques that were used in this project, complete with any assumptions or limitations that apply to them (sources: SPSS Inc, 1998, Siegel and Castellan, 1988).

- **Kruskal-Wallis Test (KW):** The KW test the nonparametric equivalent of the ANOVA, it tests distributions of scores by ranking each score and then comparing the sample mean ranks. As with the ANOVA, it is ideal for determining differences between three or more samples but it does not reveal the nature of the differences. Post-hoc tests are required to isolate specifically where the differences occur. The typical post-hoc method for the KW test compares the differences of the mean ranks (similar to the difference of proportions test), a test in which a Z-score is calculated and then compared to the standardized normal distribution. The KW test assumes a continuous distribution with data that are at least ordinal. The KW test computes a chi-square (χ^2) test statistic and compares this value to a known distribution of χ^2.

- **Chi-square Test:** The chi-square test is used to compare frequencies of data that are divided into discrete categories (i.e. nominal data). The null hypothesis is that the frequencies of data for two or more groups are the same. One step in completing the chi-square test is dividing the data into categories and grouping variables via a contingency table (see Table 7.4 for an example). The test then computes expected cell values based upon the row and column totals and the expected frequencies are compared to the observed frequencies to determine significant differences. The underlying assumptions are that no cell is can have an expected frequency less than one, and no more than twenty percent of the cells can have expected frequencies less than five. The chi-square test is another omnibus test which simply reveals that some difference does exist in the groups, post-hoc comparisons to determine the nature of these differences involve compared standardized differences between the expected and observed results to the standard normal distribution (Z). This test computes a chi-square (χ^2) test statistic and compares this value to a known distribution of χ^2.

- **Spearman Rank-order Correlation:** Spearman's Rho (?) is the nonparametric equivalent to the Pearson's Correlation Coefficient (r). It tests for a linear relationship between the rankings of the scores, however, as opposed to the scores themselves. It is interpreted in a similar manner and requires at least ordinal data.

7.2.6 Presentation of Statistical Results

Where statistical tests are presented in this document, they are presented according to the following format.

(TS(df,df) = VAL, p = PROB)

where

TS = The calculated test statistic (e.g. t, Z, or χ^2)

df = Degrees of freedom, the number and range of df's vary from test to test, some test statistics (e.g. Z) have no degrees of freedom

VAL = The value of the calculated test statistic

PROB = The significance level. When the significance level is less than the specified alpha level, the actual significance will not be presented it will just be stated that it is less than alpha

7.3 Participant Data

In this section, detailed quantitative data concerning the research subjects will be presented. First will be a short section on the instructors and following will be sections on the fall students, the spring students, and all students combined. The data presented herein will not only present important information about the study groups but will also make comparisons among the groups on demographic and non-experimental variables.

7.3.1 Instructors

Three instructors taught the six sections of CEE 2020 that were used in this experiment. The instructors were all of different rank; one was a full professor, one was a senior academic professional, and one was an instructor. Two of the instructors were male and one was a female. Two had been teaching the course for number of years, and one was teaching the course for only the second time. Anecdotally, all three of the instructors have gained reputations among students as being good teachers and are highly sought after. Finally, all three instructors had very good student evaluation records.

The lectures within the domain of truss analysis, however, along with the special session of the intervention were all taught by a guest lecturer. As detailed above, a guest lecture was used in order to eliminate any bias that may be introduced as a result of having different instructors teach the truss analysis portion of the course.

To further alleviate any fears regarding instructor bias, a simple assessment of the instructors was conducted. A popular instructor assessment technique, the Teaching Goals Inventory (Angelo and Cross, 1993), was completed by each of the instructors. One of the purposes of the inventory is to help instructors become aware of goals they wish to accomplish within an individual course. This analysis was used in this project as a means of measuring whether or not all three instructors approached the course with the same goals in mind. In the case of this assessment, each of the three instructors completed the inventory specifically for CEE 2020. The instructors completed the inventory by ranking 52 goals categorized into six clusters, as essential, very important, important, unimportant, or not applicable.

The six clusters in the inventory are as follows (Angelo and Cross, 1993, 393-397): 1). Higher order thinking skills, which includes goals such as “Develop ability to apply principles and generalizations already learned to new problems and situations” and “Develop ability to draw reasonable inferences from observations”. 2). Basic academic success skills, which includes such goals as “Improve skill at paying attention” and “Improve mathematic skills”. 3). Discipline-specific knowledge and skills, including such goals as “Learn concepts and theories in this subject” and “Learn to evaluate methods and materials in this subject”. 4). Liberal arts and academic values, which includes goals such as “Develop an informed historical perspective” and “Develop capacity to make informed ethical choices”. 5). Work and career preparation, including goals such as “Develop leadership skills” and “Develop a commitment to accurate work”. 6). Personal development, which includes such goals as “Cultivate emotional health and well-being” and “Cultivate a sense of responsibility for one’s own behavior”. The instructors were not to merely rate the merit of the goal nor were they instructed to rate the goals as outcomes they hoped for the students to gain; they were instructed to rate the importance of each goal based on what they actually strive for their students to accomplish.

The results of this assessment are summarized by the cluster scores shown in table 7.1. Cluster scores were simply calculated by finding the average of the ratings (5 being essential and 1 being not applicable) for the goals within a cluster of goals (each of which included between 8 and 10 of the 52 total goals). As can be seen, ratings varied among the instructors and Instructors A and B consistently had higher cluster scores than did instructor C. This is typical with rating type assessments, such as likert scales, in which some individuals tend to rate consistently on the lower side while others tend to consistently rate more generously. This type of situation is ideal for nonparametric statistics, which are based not upon the actual scores but upon the ranking or ordering of the scores.

As such, the rankings for each of the clusters scores for each of the professors are given in table 7.2. This table shows the cluster on which each instructor scored highest, second highest, and so on. At first glance, the rankings may seem very different. A closer look however, reveals that the only real difference in the rankings occurred in the personal development cluster, which varied among the instructors. This is understandable considering the content of the personal development cluster; the goals in this cluster are very personal and would be expected to vary widely, even among professors of the same discipline. Some instructors might sincerely strive to help students improve their self-esteem while others may not consider this an essential part of their teaching. If the personal development cluster is removed from the analysis, the rankings are now identical (with the exception of a tied cluster score for Instructor B) for each of the three instructors as shown in table 7.3.

Table 7.1 Instructors cluster scores on the Teaching Goals Inventory.

Cluster Name	Cluster Scores Instructor A	 Instructor B	 Instructor C
Higher Order Thinking Skills	4.00	4.25	3.38
Basic Academic Success Skills	2.78	2.67	2.11
Discipline Specific Knowledge and Skills	3.00	3.75	2.50
Liberal Arts and Acedemic Values	2.30	2.30	1.30
Work and Career Preparation	3.13	3.75	2.88
Personal Development	3.56	2.44	2.44

Table 7.2 Rankings of cluster scores on Teaching Goals Inventory for each instructor.

Cluster Name	Cluster Rankings Instructor A	 Instructor B	 Instructor C
Higher Order Thinking Skills	1	1	1
Basic Academic Success Skills	5	4	5
Discipline Specific Knowledge and Skills	4	2	3
Liberal Arts and Acedemic Values	6	6	6
Work and Career Preparation	3	2	2
Personal Development	2	5	4

Table 7.3 Rankings of cluster scores on Teaching Goals Inventory for each instructor with the personal development cluster excluded.

Cluster Name	Cluster Rankings Excluding Personal Development: Instructor A	Instructor B	Instructor C
Higher Order Thinking Skills	1	1	1
Basic Academic Success Skills	4	4	4
Discipline Specific Knowledge and Skills	3	2	3
Liberal Arts and Acedemic Values	5	5	5
Work and Career Preparation	2	2	2

A nonparametric analysis was performed on the cluster scores to determine if the distributions were similar. A bivariate correlation using Spearman's rho revealed that a significant correlation existed between distribution of scores for Instructor C and Instructor B (?=.928, p<.05) and between Instructor C and Instructor A (?=.829, p<.05). No significant correlation existed between the distribution of scores for Instructors A and B however. Also, though two significant relationships were discovered, when the significance level is adjusted for a familywise error rate via Bonferroni, the pairwise significance becomes .0167, a level at which the correlation between Instructor C's scores and Instructor A's scores is no longer significant.

If, however, the personal development cluster is not considered in the analysis, the results are quite different; a significant relationship existed between all pairs at the familywise error rate of .05. The correlation between scores for Instructors A and B was significant (?=.975, p<.01), as was the relationship between Instructors B and C (?=.975, p<.01) and the relationship between Instructors C and A (?=1.0, p<.01). A 1.0

correlation coefficient between the distribution of scores for Instructors C and A implies that they are identical when in fact it is the rankings of the scores that are identical, as shown in table 7.3.

As a side note, this is an excellent illustration into how such nonparametric statistics are calculated. It is important to note that the nonparametric correlation analysis was performed on the scores (table 7.1) and not the rankings (tables 7.2 and 7.3). It is clear from the results, however, that part of the analysis is the ranking of the scores prior to comparison and that the ultimate comparison takes place between the rankings and not the scores. In other words, though the rankings were explicitly presented in tables 7.2 and 7.3, this is not normally done and was performed here just for illustration; the actual analysis is performed on the scores and the ranking of the scores is an integral, internal step in the nonparametric analysis.

All of this was done to provide further evidence against the possibility of an instructor-related bias. Whether a bias actually existed or not cannot be assessed. Through the use of the Teaching Goals Inventory and by implementing a guest lecturer, sufficient attempts were made to ensure that the results of the assessments would not be tainted by the fact that the subject groups had different instructors.

7.3.2 Student Participants

Six sections of CEE 2020 with a total of 281 students participated in this project. This section will present detailed information about these students and the research groups to which they belong. The section will be divided by semesters and will then look at both semesters combined.

7.3.2.1 Fall 2002

As stated previously, three sections of CEE 2020 from the fall semester of 2002 participated in the study.

Section C met from 10:05 A.M. to 10:55 A.M on Mondays, Wednesdays, and Fridays and was taught by Instructor A. Section C was assigned to the content-type experimental group and used the educational tool Multimedia Engineering Statics during the special session of the intervention. Section C included 45 students, four of who withdrew from the class at some point in time during the semester. Of the remaining students, 19 were industrial engineering majors, 15 were civil engineering students, four majored in electrical engineering, two were textile and fiber engineering majors, and one was a computer engineering student. Twenty-two of the participants were male and 19 were female.

Thirty-seven students completed the retention posttest at which time the following demographic data were collected. Twelve of the students were sophomores, 19 were juniors, and six were seniors. Ten students had a grade point average (GPA) in the 3.5-4.0 range, 11 students had GPAs ranging from 3.0-3.5, ten students were in the 2.5-3.0 range, and five had a GPA between 2.0 and 2.5. A majority of the students took between 12 and 17 credits in Fall 2002, but one student was registered for 18 credits and two students were taking 10 credits. Twenty-six of the students were Caucasian, six were Asian/Pacific Islanders, two were African Americans, one was Hispanic, and two were Indian.

Section E met from 12:05 A.M. to 12:55 A.M on Mondays, Wednesdays, and Fridays and was taught by Instructor A. Section E was assigned to the tool-type experimental group and used the structural analysis and design software Dr. Frame during the special session of the intervention. Section E included 42 students, four of who withdrew from the class at some point in time during the semester. Of the remaining students, 17 were industrial engineering majors, 13 were civil engineering students, six majored in electrical engineering, and two were textile and fiber engineering majors. Nineteen of the participants were male and 19 were female.

Thirty-seven students completed the retention posttest at which time the following demographic data were collected. Twelve of the students were sophomores, 22 were juniors, and three were seniors. Six students had a GPA in the 3.5-4.0 range, 12 students had GPAs ranging from 3.0-3.5, 12 students were in the 2.5-3.0 range, five had a GPA between 2.0 and 2.5, and two students had GPAs below 2.0. A majority of the students took between 12 and 17 credits in Fall 2002, but one student was registered for 19 credits, one was taking eleven credits, one was registered for nine credits, and one was taking only this class. Twenty-eight of the students were Caucasian, four were Asian/Pacific Islanders, one was African American, two were Hispanic, and two were Indian.

Section H met from 8:05 A.M. to 9:25 A.M on Tuesdays and Thursdays and was taught by Instructor B. Section H was assigned to the comparison/control group and used the no software during the special session of the intervention. Section H included 71 students, one of whom withdrew from the class at some point in time during the semester. Of the remaining students, 29 were industrial engineering majors, 28 were civil engineering students, eleven majored in electrical engineering, one was a management major, and one majored in computer engineering. Forty of the participants were male and 30 were female.

Sixty-three students completed the retention posttest at which time the following demographic data were collected. Eleven of the students were sophomores, 17 were juniors, and thirty-five were seniors. Twenty-two students had a GPA in the 3.5-4.0 range, 18 students had GPA s ranging from 3.0-3.5, 17 students were in the 2.5-3.0 range, five had a GPA between 2.0 and 2.5, and one student had a GPA below 2.0. A majority of the students took between 12 and 17 credits in Fall 2002, but one student was registered for 19 credits, one was taking 21 credits, and eleven students reported taking less than 12 credits that semester. Forty-four of the students were Caucasian, seven were Asian/Pacific Islanders, five were African Americans, and six were Hispanic.

Comparisons of these data were made among the groups in an effort to show the groups were similar despite the lack of random selection or sampling. The first comparison, shown in Table 7.4, shows the breakdown of each of the section's population by different majors (note that these numbers differ slightly from those given above because they include the students who withdrew from the course). For the purpose of this comparison, small categories with small frequencies had to be lumped together in order to meet the assumptions of the chi-square test. As such, the Other Majors category is mostly made up of electrical engineering majors but also includes some textile and fiber and computer engineering students and some management students. Note that in each case, the percentage of IE students was in the low-to-mid forties and the other majors were fairly well clustered around 20 percent. There was a small difference in the percentages of civil engineering students which ranged from 33 to 41 percent. A chi-square test, however, revealed that these differences were not significant ($?^2(4)=0.724$, p=.948). Thus, the hypothesis that these three sections have similar distributions of majors cannot be rejected and any differences in the frequencies are assumed to be due to chance.

Table 7.4 Frequencies of Students' Majors per Section of CEE 2020, Fall 2002

			Major			
			Other Majors	Industrial Engineering	Civil Engineering	Total
Section ID	Section C, Fall 2002	Count	8	20	17	45
		% within Section ID	17.8%	44.4%	37.8%	100.0%
	Section E, Fall 2002	Count	9	19	14	42
		% within Section ID	21.4%	45.2%	33.3%	100.0%
	Section H, Fall 2002	Count	13	29	29	71
		% within Section ID	18.3%	40.8%	40.8%	100.0%
Total		Count	30	68	60	158
		% within Section ID	19.0%	43.0%	38.0%	100.0%

A similar comparison was done for student rank. The frequencies of student rank per section are presented in Table 7.5. The distribution of students by rank is clearly different in Section H than in the other two sections. The reason for this has been mentioned previously. During the initial phases of registration, Sections C and E are only open to sophomores, whereas Section H is open to all students and rapidly fills up with seniors who are eligible to register before juniors or sophomores are. A chi-square test reveals that the difference is significant (?^{2}(4)=30.096, $p<.01$) and so the sections cannot be assumed to have similar populations in terms of rank.

While this is not ideal, it does not greatly affect the research or the results. Rather, this is an excellent example of the why replication is required in a quasi-experimental research design. This is illustrated by the fact that in one semester, as shown above, the sections were not equal in terms of student rank. If this study were completed only once and an effect was realized, there would be no way of knowing whether the effect was a result of the different treatments that the sections received or if the result was an effect of the sample groups being made up of significantly different students. Replication, however, allows the experiment to be completed again with different sample groups and subjects. As will be seen in the next section, in the spring semester of 2003 there were no significant differences among three groups in terms of student rank. Thus, if an effect is realized in both the original and the replicated study, then it could be concluded that the effect was due to the intervention and not a result of sample bias.

Table 7.5 Frequencies of Student Rank per Section of CEE 2020, Fall 2002.

			Rank			Total
			Sophomore	Junior	Senior	
Section ID	Section C, Fall 2002	Count	12	19	6	37
		% within Section ID	32.4%	51.4%	16.2%	100.0%
	Section E, Fall 2002	Count	12	22	3	37
		% within Section ID	32.4%	59.5%	8.1%	100.0%
	Section H, Fall 2002	Count	11	17	35	63
		% within Section ID	17.5%	27.0%	55.6%	100.0%
Total		Count	35	58	44	137
		% within Section ID	25.5%	42.3%	32.1%	100.0%

Table 7.6 contains gender data per each section. One of the sections has the same number of females as males, while the other two sections have a few more males than there are females. As a side note, the percentage of females in each of the sections is unusually high. The percentages do not, however, differ significantly between the three sections ($?^2(2)=0.460$, p=.795).

Table 7.6 Breakdown of Students in CEE 2020 by Gender, Fall 2002

			Gender		
			Male	Female	Total
Section ID	Section C, Fall 2002	Count	25	20	45
		% within Section ID	55.6%	44.4%	100.0%
	Section E, Fall 2002	Count	21	21	42
		% within Section ID	50.0%	50.0%	100.0%
	Section H, Fall 2002	Count	40	31	71
		% within Section ID	56.3%	43.7%	100.0%
Total		Count	86	72	158
		% within Section ID	54.4%	45.6%	100.0%

Sections were compared for GPA as well; the data for this comparison are in Table 7.7. The percentages are similar in the three upper ranges, with all of the percentages in these ranges being about 30%, with the exception of Section E, which only had 16.2% of the students in the 3.5-4.0 range. This difference, however, did not prove to be significant when a chi-square test was performed on the data ($?^2(6)=4.863$, $p=.562$). Thus, any differences can be assumed to be a result of chance and the groups should be considered to have similar distributions of GPA.

The sections were also compared to see if there were differences in the number of credits that students were taking during Fall 2002. These data are presented in Table 7.8. Again, in order to accurately calculate the chi-square statistic, some of the categories had to be combined, which is why the lower category includes all students taking less than 13 credits and the highest category includes all students taking over 15 credits. Sections C and E have similar distributions, but section H has a greater percentage of students taking less than 13 credits. A chi-square test, however, revealed that the distributions did not differ significantly (?^{2}(8)=8.794, p=.360).

Table 7.7 Frequencies of Student GPA per Section of CEE 2020, Fall 2002

			GPA				
			<2.499	2.5-2.999	3.0-3.499	3.5-4.000	Total
Section ID	Section C, Fall 2002	Count	5	10	11	10	36
		% within Section ID	13.9%	27.8%	30.6%	27.8%	100.0%
	Section E, Fall 2002	Count	7	12	12	6	37
		% within Section ID	18.9%	32.4%	32.4%	16.2%	100.0%
	Section H, Fall 2002	Count	6	17	18	22	63
		% within Section ID	9.5%	27.0%	28.6%	34.9%	100.0%
Total		Count	18	39	41	38	136
		% within Section ID	13.2%	28.7%	30.1%	27.9%	100.0%

The final comparison was conducted to see if the students' ethnicities differed between sections. These data are presented in Table 7.9. Because so few minorities were registered for each of the sections, however, all of the minorities had to be lumped into one category in order to correctly calculate the chi-square statistic. In each of the three sections, 70 to 75 percent of the students were Caucasian and the sections did not differ significantly in terms of ethnicity (?^{2}(2)=0.336, p=.845).

In summary, only with regards to the student rank variable did the sample groups differ significantly, in all other measured demographic areas, the sections can be assumed to be equal. A similar analysis of the spring data will now be presented.

Table 7.8 Frequencies of Credits Taken by Students of CEE 2020 During Fall, 2002

			Credits					
			<13	13	14	15	>15	Total
Section ID	Section C, Fall 2002	Count	6	6	7	7	11	37
		% within Section ID	16.2%	16.2%	18.9%	18.9%	29.7%	100.0%
	Section E, Fall 2002	Count	5	6	7	8	10	36
		% within Section ID	13.9%	16.7%	19.4%	22.2%	27.8%	100.0%
	Section H, Fall 2002	Count	20	13	6	7	16	62
		% within Section ID	32.3%	21.0%	9.7%	11.3%	25.8%	100.0%
Total		Count	31	25	20	22	37	135
		% within Section ID	23.0%	18.5%	14.8%	16.3%	27.4%	100.0%

Table 7.9 Breakdown of CEE 2020 Students by Ethnicity, Fall 2002

			Ethnicity		
			Other	Caucasion	Total
Section ID	Section C, Fall 2002	Count	11	26	37
		% within Section ID	29.7%	70.3%	100.0%
	Section E, Fall 2002	Count	9	28	37
		% within Section ID	24.3%	75.7%	100.0%
	Section H, Fall 2002	Count	18	44	62
		% within Section ID	29.0%	71.0%	100.0%
Total		Count	38	98	136
		% within Section ID	27.9%	72.1%	100.0%

7.3.2.2 Spring 2003

Section B met from 9:05 A.M. to 9:55 A.M. on Mondays, Wednesdays, and Fridays and was taught by Instructor C. Section B was assigned to the content-type experimental group and used Multimedia Engineering Statics during the special session of the intervention. Section B included 29 students, of whom 11 were industrial engineering majors, 16 were civil engineering students, one majored in electrical engineering, and one was a management student. Nineteen of the participants were male and ten were female.

Twenty-six students completed the retention posttest at which time the following demographic data were collected. Nine of the students were sophomores, five were juniors, and twelve were seniors. One student had a GPA in the 3.5-4.0 range, eight students had GPAs ranging from 3.0-3.5, ten students were in the 2.5-3.0 range, five had a GPA between 2.0 and 2.5, and one student had a GPA below 2.0. A majority of the students took between 12 and 17 credits in Fall 2002, but one student was registered for 19 credits, two were taking 18 credits, and two were taking eleven credits.   Twenty-one of the students were Caucasian, two were Asian/Pacific Islanders, one was African American, and two were Hispanic.

Section E met from 12:05 P.M. to 12:55 P.M on Mondays, Wednesdays, and Fridays and was taught by Instructor A. Section E was assigned to the Tool-type experimental group and used the structural analysis and design software Dr. Frame during the special session of the intervention. Section E included 46 students, two of who withdrew from the class at some point in time during the semester. Of the remaining students, 17 were industrial engineering majors, 21 were civil engineering students, three majored in electrical engineering, two were textile and fiber engineering majors and one was a computer engineering student. Twenty-three of the participants were male and 21 were female.

Forty-four students completed the retention posttest at which time the following demographic data were collected. Twenty of the students were sophomores, 14 were juniors, and ten were seniors. Thirteen students had a GPA in the 3.5-4.0 range, 12 students had GPAs ranging from 3.0-3.5, ten students were in the 2.5-3.0 range, four had a GPA between 2.0 and 2.5, and four students had GPAs below 2.0. A majority of the students took between 12 and 17 credits in Fall 2002, but one student was registered for only six credits. Thirty-three of the students were Caucasian, five were Asian/Pacific Islanders, three were African Americans, and one was Hispanic.

Section G met from 2:05 P.M. to 2:55 P.M on Mondays, Wednesdays, and Fridays and was taught by Instructor A. Section G was assigned to the comparison/control group and used no software during the special session of the intervention. Section G included 48 students, three of whom withdrew from the class at some point in time during the semester. Of the remaining students, 26 were industrial engineering majors, nine were civil engineering students, eight majored in electrical engineering, one was a textile and fiber engineering majors and one was a computer engineering student. Thirty-one of the participants were male and 14 were female.

Thirty-seven students completed the retention posttest at which time the following demographic data were collected. Sixteen of the students were sophomores, nine were juniors, and 12 were seniors. Seven students had a GPA in the 3.5-4.0 range, ten students had GPAs ranging from 3.0-3.5, six students were in the 2.5-3.0 range, nine had a GPA between 2.0 and 2.5, and one student had a GPA below 2.0. A majority of the students took between 12 and 17 credits in Fall 2002, but one student was registered for 20 credits, two were registered for 21 credits, and fiver were taking less than 11 credits. Twenty-four of the students were Caucasian, six were Asian/Pacific Islanders, four were African Americans, two were Hispanic, and two were Indian.

The same comparisons that were made for the fall data were also made for the spring data. The first comparison was for student majors; see Table 7.10 for this data. As the data in Table 7.10 show, Section G had a much higher percentage of IE students than the other two sections. A chi-square test revealed that this difference was significant (?^{2}(4)=12.863, p<.05) and thus the populations cannot be assumed to be equal in terms of majors. As mentioned previously however, the fall data did not reveal a significant difference in the subject groups in terms of majors and so if an effect is revealed in both studies it can be assumed to not be a result of differences in student majors.

Student rank data from Spring 2003 are presented in Table 7.11. Despite there being some differences between the frequency distributions in the three sections, a chi-square test revealed that these differences are not significant (?^{2}(4)=4.369, p=.358) and thus can be assumed to be a chance result.

No significant differences were revealed for gender either (?^{2}(2)=2.402, p=.301). The gender data are presented in table 7.12. Males were the majority in each section, but there were considerable numbers of females in each section as well.

Table 7.10 Frequencies of Students' Majors per Section of CEE 2020, Spring 2003

			Major			
			Other Majors	Industrial Engineering	Civil Engineering	Total
Section ID	Section B, Spring 2003	Count	2	11	16	29
		% within Section ID	6.9%	37.9%	55.2%	100.0%
	Section E, Spring 2003	Count	6	19	21	46
		% within Section ID	13.0%	41.3%	45.7%	100.0%
	Section G, Spring 2003	Count	10	29	9	48
		% within Section ID	20.8%	60.4%	18.8%	100.0%
Total		Count	18	59	46	123
		% within Section ID	14.6%	48.0%	37.4%	100.0%

Table 7.11 Frequencies of Student Rank per Section of CEE 2020, Spring 2003

			Rank			Total
			Sophomore	Junior	Senior	
Section ID	Section B, Spring 2003	Count	9	5	12	26
		% within Section ID	34.6%	19.2%	46.2%	100.0%
	Section E, Spring 2003	Count	20	14	10	44
		% within Section ID	45.5%	31.8%	22.7%	100.0%
	Section G, Spring 2003	Count	16	9	12	37
		% within Section ID	43.2%	24.3%	32.4%	100.0%
Total		Count	45	28	34	107
		% within Section ID	42.1%	26.2%	31.8%	100.0%

Table 7.12 Breakdown of Students in CEE 2020 by Gender, Spring 2003.

			Gender		Total
			Male	Female	
Section ID	Section B, Spring 2003	Count	19	10	29
		% within Section ID	65.5%	34.5%	100.0%
	Section E, Spring 2003	Count	24	22	46
		% within Section ID	52.2%	47.8%	100.0%
	Section G, Spring 2003	Count	32	16	48
		% within Section ID	66.7%	33.3%	100.0%
Total		Count	75	48	123
		% within Section ID	61.0%	39.0%	100.0%

Table 7.13 presents GPA data for the sections that participated in Spring 2003. Again, all GPA counts below 2.5 were lumped together in order to meet the assumptions of the chi-square test. The test revealed that there was no significant difference between the different distributions of GPA between the three sections ($?^2(6)=9.218$, p=.162).

Table 7.13 Distributions of Student GPA per Section of CEE 2020, Spring 2003

			GPA				
			<2.499	2.5-2.999	3.0-3.499	3.5-4.000	Total
Section ID	Section B, Spring 2003	Count	6	10	8	1	25
		% within Section ID	24.0%	40.0%	32.0%	4.0%	100.0%
	Section E, Spring 2003	Count	8	10	12	13	43
		% within Section ID	18.6%	23.3%	27.9%	30.2%	100.0%
	Section G, Spring 2003	Count	10	6	10	7	33
		% within Section ID	30.3%	18.2%	30.3%	21.2%	100.0%
Total		Count	24	26	30	21	101
		% within Section ID	23.8%	25.7%	29.7%	20.8%	100.0%

A similar comparison was made with the number of credits that students were taking during the spring semester of 2003. These data are presented in Table 7.14. A combination of categories, identical to what was done with the fall data as described above, was required with these data as well. While there are some observable differences between the different frequencies, these differences are not significant ($?^2(8)=5.768$, $p=.673$) and thus the sections can be assumed to be similar with regards to the number of credits that students were taking.

Table 7.14 Frequencies of Students Credits per Section of CEE 2020, Spring 2003

			Credits					
			<13	13	14	15	>15	Total
Section ID	Section B, Spring 2003	Count	8	5	2	6	5	26
		% within Section ID	30.8%	19.2%	7.7%	23.1%	19.2%	100.0%
	Section E, Spring 2003	Count	11	7	11	7	8	44
		% within Section ID	25.0%	15.9%	25.0%	15.9%	18.2%	100.0%
	Section G, Spring 2003	Count	10	3	6	9	9	37
		% within Section ID	27.0%	8.1%	16.2%	24.3%	24.3%	100.0%
Total		Count	29	15	19	22	22	107
		% within Section ID	27.1%	14.0%	17.8%	20.6%	20.6%	100.0%

Again, the final comparison was for ethnicity and these data are shown in table 7.15. As with the fall data, each of the sections were predominately Caucasian and all the minorities were lumped into one category, the *other* category, in order to meet the requirements of the chi-square test. This test revealed no significant difference between the three subject groups in terms of ethnicity ($?^2(2)=3.279$, $p=.194$).

In summary, the only demographic variable in which a significant difference occurred between the three different subject groups was the students' major variable. Because there was no difference in this variable in the fall study, replicated results could overcome the effects of this sample bias.

Table 7.15 Breakdown of CEE 2020 Students by Ethnicity, Spring 2003

			Ethnicity		
			Other	Caucasion	Total
Section ID	Section B, Spring 2003	Count	5	21	26
		% within Section ID	19.2%	80.8%	100.0%
	Section E, Spring 2003	Count	11	33	44
		% within Section ID	25.0%	75.0%	100.0%
	Section G, Spring 2003	Count	14	22	36
		% within Section ID	38.9%	61.1%	100.0%
Total		Count	30	76	106
		% within Section ID	28.3%	71.7%	100.0%

7.3.2.3 Combined Data - All Students

Data from all six groups (i.e. both semesters) are presented in this section in order to describe all the participants that were involved in the study. As with the previous sections, these data will be presented in a form that is cross-tabulated by section for two reasons: first, these tables succinctly present the frequencies of the demographic and non-experimental variables and second, the cross-tabulation is a necessary step in the calculation of the chi-square statistic.

Table 7.16 presents data on students' major of study for all six semesters that participated. With the exception of Section G, Spring 2003, the distributions are fairly similar. When all six sections are looked at together, any differences that are present are no longer significant (?^{2}(10)=14.478, p=.152). The data were further analyzed by combining sections that underwent the same treatment conditions. Sections C from fall and B from spring were combined because they were both exposed to the content-type treatment condition. Similarly, Sections E from fall and E from spring were combined based on both being in the tool-type treatment condition. Finally, Sections H from fall and G from spring were combined into the comparison group. The data from both semesters tabulated by treatment condition are in Table 7.17. The frequencies are more similarly distributed when the sections are combined and the chi-square test revealed no significant differences (?^{2}(4)=3.573, p=.467).

Table 7.16 Major of Study by Section, All Students

			Major			Total
			Other Majors	Industrial Engineering	Civil Engineering	
Section ID	Section C, Fall 2002	Count	8	20	17	45
		% within Section ID	17.8%	44.4%	37.8%	100.0%
	Section E, Fall 2002	Count	9	19	14	42
		% within Section ID	21.4%	45.2%	33.3%	100.0%
	Section H, Fall 2002	Count	13	29	29	71
		% within Section ID	18.3%	40.8%	40.8%	100.0%
	Section B, Spring 2003	Count	2	11	16	29
		% within Section ID	6.9%	37.9%	55.2%	100.0%
	Section E, Spring 2003	Count	6	19	21	46
		% within Section ID	13.0%	41.3%	45.7%	100.0%
	Section G, Spring 2003	Count	10	29	9	48
		% within Section ID	20.8%	60.4%	18.8%	100.0%
Total		Count	48	127	106	281
		% within Section ID	17.1%	45.2%	37.7%	100.0%

Table 7.17 Major of Study by Treatment Condition, All Students

			Major			Total
			Other Majors	Industrial Engineering	Civil Engineering	
Treatment Condition	Content-type Software Group	Count	10	31	33	74
		% within Treatment Condition	13.5%	41.9%	44.6%	100.0%
	Tool-type Software Group	Count	15	38	35	88
		% within Treatment Condition	17.0%	43.2%	39.8%	100.0%
	Comparison Group	Count	23	58	38	119
		% within Treatment Condition	19.3%	48.7%	31.9%	100.0%
Total		Count	48	127	106	281
		% within Treatment Condition	17.1%	45.2%	37.7%	100.0%

Similar comparisons were completed for the rank variable. Table 7.18 presents rank data for each of the six sections that participated in the study. The differences that occurred between the fall sections, as previously discussed, were significant, even when compared to all the other participating sections (?^{2}(10)=44.559, p<.01). When the sections were combined and compared based upon treatment condition, as shown in Table 7.19, the differences were still significant (?^{2}(4)=20.374, p<.01). This is further evidence for the need of replication. As mentioned previously, the spring data did not differ significantly in terms of rank and so conducting the study in both spring and fall, and not merely looking at the aggregate of the two, will eliminate a possible sample bias for the rank variable.

Table 7.18 Student Rank by Section, All Students

			Rank			Total
			Sophomore	Junior	Senior	
Section ID	Section C, Fall 2002	Count	12	19	6	37
		% within Section ID	32.4%	51.4%	16.2%	100.0%
	Section E, Fall 2002	Count	12	22	3	37
		% within Section ID	32.4%	59.5%	8.1%	100.0%
	Section H, Fall 2002	Count	11	17	35	63
		% within Section ID	17.5%	27.0%	55.6%	100.0%
	Section B, Spring 2003	Count	9	5	12	26
		% within Section ID	34.6%	19.2%	46.2%	100.0%
	Section E, Spring 2003	Count	20	14	10	44
		% within Section ID	45.5%	31.8%	22.7%	100.0%
	Section G, Spring 2003	Count	16	9	12	37
		% within Section ID	43.2%	24.3%	32.4%	100.0%
Total		Count	80	86	78	244
		% within Section ID	32.8%	35.2%	32.0%	100.0%

Table 7.19 Student Rank by Treatment Condition, All Students

			Rank			
			Sophomore	Junior	Senior	Total
Treatment Condition	Content-type Software Group	Count	21	24	18	63
		% within Treatment Condition	33.3%	38.1%	28.6%	100.0%
	Tool-type Software Group	Count	32	36	13	81
		% within Treatment Condition	39.5%	44.4%	16.0%	100.0%
	Comparison Group	Count	27	26	47	100
		% within Treatment Condition	27.0%	26.0%	47.0%	100.0%
Total		Count	80	86	78	244
		% within Treatment Condition	32.8%	35.2%	32.0%	100.0%

There were no significant differences in gender between sections within semesters as described above. When the data were compared across semesters, as shown in Table 7.20, there were still no significant differences (?^{2}(5)=4.103, p=.548). When sections that underwent similar treatment conditions were combined and compared, see Table 7.21, there were no significant differences between the groups in terms of gender either (?^{2}(2)=2.007, p=.367).

Table 7.20 Gender by Section, All Students

			Gender		
			Male	Female	Total
Section ID	Section C, Fall 2002	Count	25	20	45
		% within Section ID	55.6%	44.4%	100.0%
	Section E, Fall 2002	Count	21	21	42
		% within Section ID	50.0%	50.0%	100.0%
	Section H, Fall 2002	Count	40	31	71
		% within Section ID	56.3%	43.7%	100.0%
	Section B, Spring 2003	Count	19	10	29
		% within Section ID	65.5%	34.5%	100.0%
	Section E, Spring 2003	Count	24	22	46
		% within Section ID	52.2%	47.8%	100.0%
	Section G, Spring 2003	Count	32	16	48
		% within Section ID	66.7%	33.3%	100.0%
Total		Count	161	120	281
		% within Section ID	57.3%	42.7%	100.0%

Table 7.21 Gender by Treatment Condition, All Students

			Gender		
			Male	Female	Total
Treatment Condition	Content-type Software Group	Count	44	30	74
		% within Treatment Condition	59.5%	40.5%	100.0%
	Tool-type Software Group	Count	45	43	88
		% within Treatment Condition	51.1%	48.9%	100.0%
	Comparison Group	Count	72	47	119
		% within Treatment Condition	60.5%	39.5%	100.0%
Total		Count	161	120	281
		% within Treatment Condition	57.3%	42.7%	100.0%

Similar results occurred for GPA, credits, and ethnicity as shown in Tables 7.22 through 7.27. None of these comparisons was significant between all sections: GPA ($?^2(15)=18.383$, p=.243), credits ($?^2(20)=17.266$, p=.636), ethnicity ($?^2(5)=3.632$, p=.604). There were no significant differences for treatment conditions either: GPA (?[2](6)=3.505, p=.743), credits (?[2](14)=13.244, p=.507), ethnicity (?[2](2)=1.699, p=.428). Thus, for these variables, it can be assumed that the groups are equal whether considering individual semesters or aggregate data from combining semesters.

Table 7.22 GPA by Section, All Students

			GPA				Total
			2.0-2.499	2.5-2.999	3.0-3.499	3.5-4.000	
Section ID	Section C, Fall 2002	Count	5	10	11	10	36
		% within Section ID	13.9%	27.8%	30.6%	27.8%	100.0%
	Section E, Fall 2002	Count	7	12	12	6	37
		% within Section ID	18.9%	32.4%	32.4%	16.2%	100.0%
	Section H, Fall 2002	Count	6	17	18	22	63
		% within Section ID	9.5%	27.0%	28.6%	34.9%	100.0%
	Section B, Spring 2003	Count	6	10	8	1	25
		% within Section ID	24.0%	40.0%	32.0%	4.0%	100.0%
	Section E, Spring 2003	Count	8	10	12	13	43
		% within Section ID	18.6%	23.3%	27.9%	30.2%	100.0%
	Section G, Spring 2003	Count	10	6	10	7	33
		% within Section ID	30.3%	18.2%	30.3%	21.2%	100.0%
Total		Count	42	65	71	59	237
		% within Section ID	17.7%	27.4%	30.0%	24.9%	100.0%

Table 7.23 GPA by Treatment Condition, All Stude nts

			GPA				Total
			2.0-2.499	2.5-2.999	3.0-3.499	3.5-4.000	
Treatment Condition	Content-type Software Group	Count	11	20	19	11	61
		% within Treatment Condition	18.0%	32.8%	31.1%	18.0%	100.0%
	Tool-type Software Group	Count	15	22	24	19	80
		% within Treatment Condition	18.8%	27.5%	30.0%	23.8%	100.0%
	Comparison Group	Count	16	23	28	29	96
		% within Treatment Condition	16.7%	24.0%	29.2%	30.2%	100.0%
Total		Count	42	65	71	59	237
		% within Treatment Condition	17.7%	27.4%	30.0%	24.9%	100.0%

Table 7.24 Credits by Section, All Students

			Credits					
			<13	13	14	15	>15	Total
Section ID	Section C, Fall 2002	Count	6	6	7	7	11	37
		% within Section ID	16.2%	16.2%	18.9%	18.9%	29.7%	100.0%
	Section E, Fall 2002	Count	5	6	7	8	10	36
		% within Section ID	13.9%	16.7%	19.4%	22.2%	27.8%	100.0%
	Section H, Fall 2002	Count	20	13	6	7	16	62
		% within Section ID	32.3%	21.0%	9.7%	11.3%	25.8%	100.0%
	Section B, Spring 2003	Count	8	5	2	6	5	26
		% within Section ID	30.8%	19.2%	7.7%	23.1%	19.2%	100.0%
	Section E, Spring 2003	Count	11	7	11	7	8	44
		% within Section ID	25.0%	15.9%	25.0%	15.9%	18.2%	100.0%
	Section G, Spring 2003	Count	10	3	6	9	9	37
		% within Section ID	27.0%	8.1%	16.2%	24.3%	24.3%	100.0%
Total		Count	60	40	39	44	59	242
		% within Section ID	24.8%	16.5%	16.1%	18.2%	24.4%	100.0%

Table 7.25 Credits by Treatment Condition, All Students

			Credits					
			<13	13	14	15	>15	Total
Treatment Condition	Content-type Software Group	Count	14	11	9	13	16	63
		% within Treatment Condition	22.2%	17.5%	14.3%	20.6%	25.4%	100.0%
	Tool-type Software Group	Count	16	13	18	15	18	80
		% within Treatment Condition	20.0%	16.3%	22.5%	18.8%	22.5%	100.0%
	Comparison Group	Count	30	16	12	16	25	99
		% within Treatment Condition	30.3%	16.2%	12.1%	16.2%	25.3%	100.0%
Total		Count	60	40	39	44	59	242
		% within Treatment Condition	24.8%	16.5%	16.1%	18.2%	24.4%	100.0%

Table 7.26 Ethnicity by Section, All Students

			Ethnicity		
			Other	Caucasion	Total
Section ID	Section C, Fall 2002	Count	11	26	37
		% within Section ID	29.7%	70.3%	100.0%
	Section E, Fall 2002	Count	9	28	37
		% within Section ID	24.3%	75.7%	100.0%
	Section H, Fall 2002	Count	18	44	62
		% within Section ID	29.0%	71.0%	100.0%
	Section B, Spring 2003	Count	5	21	26
		% within Section ID	19.2%	80.8%	100.0%
	Section E, Spring 2003	Count	11	33	44
		% within Section ID	25.0%	75.0%	100.0%
	Section G, Spring 2003	Count	14	22	36
		% within Section ID	38.9%	61.1%	100.0%
Total		Count	68	174	242
		% within Section ID	28.1%	71.9%	100.0%

Table 7.27 Ethnicity by Treatment Condition, All Students

			Ethnicity		
			Other	Caucasion	Total
Treatment Condition	Content-type Software Group	Count	16	47	63
		% within Treatment Condition	25.4%	74.6%	100.0%
	Tool-type Software Group	Count	20	61	81
		% within Treatment Condition	24.7%	75.3%	100.0%
	Comparison Group	Count	32	66	98
		% within Treatment Condition	32.7%	67.3%	100.0%
Total		Count	68	174	242
		% within Treatment Condition	28.1%	71.9%	100.0%

7.3.2.4 Summary of Student Data

A great deal of data was gathered from the students in the six sections that participated in this project. This section has described certain demographic and non-experimental information about the students in each of the sections. In addition, comparisons were made between the different participating sections in an effort to show that the sections were similar.

It can be assumed from these comparisons that the different sections and groups of students are similar in all respects except for the following. The spring sections of CEE 2020 were significantly different in terms of major but there was no significant difference in major in the fall semester. A similar situation occurred with the rank variable except that it was the fall sections that were significantly different. While these differences in rank were not significant for the spring semesters, significant differences were still present in aggregate data when all six sections were compared. This all points to the need for replication; a necessity that was accounted for in the design of the research.

One other point of information that stood out from the student data was the high percentage of female students in each of the sections, ranging from 33 to 50 percent. This is unusual because enrollment and graduation rates for female engineering students tend to be around 20 percent. While this is unusual, and unexplainable by the data gathered in this study, it is not of concern because the female rates were high for all sections involved in the study. Indeed, there were no significant differences between any of the sections in terms of gender.

7.4 Pretest Data

As mentioned previously, the pretest was originally designed to include questions about trusses as well as questions to test students' knowledge of information required to solve trusses. The truss questions were to be used in conjunction with the pretest to obtain differential scores of learning. The prior-knowledge questions were to be used to compare the different groups involved in the study to show that they were all similar in terms of prior knowledge; that they were all starting the truss portion of the course on the same foot. During the formative assessment, however, it was determined that this combined assessment was rather lengthy. Furthermore, it was decided during the planning stages of the actual intervention, based on input from the instructors in whose classes the intervention would be taking place, that the length of the combined assessment would be a detriment to its completion. As such, the truss questions were excluded from the pretest. The purpose of the pretest, then, became singular: to assess students' knowledge prior to the truss portion of the course. As a result, the learning variable would no longer be a differential variable but would merely be assessed by comparing results from the posttest.

The final form of the pretest, containing only prior-knowledge questions, is presented in Appendix D. Scores from the first question, however, were not included in the final pretest results nor were they a part of the following analyses. This decision was made prior to analyzing the data and was based upon information gathered while the pretests were being graded. The first question involves the analysis a simple truss that requires no formal truss training to solve; the question was designed to be solved using simple vector algebra. The concern was not with the solution method, however, it was with the required solution itself. The question asked students to solve for the internal forces in the two truss members and the issue is that prior to the truss portion of the course, the phrase *forces in members* was not used. Thus, students were asked to do something they had never done, find forces internal to members, despite the fact that they should have been capable of performing the task. Because many students were not sure as to what exactly was being asked of them, many solved for the external reaction forces, which incidentally required the same skills as needed to solve for the internal forces and was a task that students had completed a number of times. This situation did not arise during the formative assessment because during that study, students had received one lecture on trusses before completing the pretest and thus knew what the phrase *forces in members* meant. Had the situation arisen during the formative phase, this question would not have been on the pretest. Because of this confusion over what was being asked of the students, the scores from this question were not included in the analysis or results of the pretest.

As described in the previous chapter, the pretest was given as a take home exam and students were allowed to take as much time as needed to complete the exam but they were asked to record the amount of time they took to complete it. The pretest was graded by the researcher who developed and adhered to a strict grading rubric to ensure that each student from each section was grading in a similar manner.

The results from these pretests are presented in the following sections beginning with the Fall 2002 data, followed by the Spring 2003 data, and concluding with a look at all the data from both semesters.

7.4.1 Fall 2002 Pretest Data

One hundred and forty-two students from the fall sections of CEE 2020 completed the pretest; 39 of these were from Section C, 38 were from Section E, and 65 were from Section H. This translates into respective completion rates of 95, 100, and 93 percents. Three difference of proportions tests revealed no significant differences between any of the return rates (Z_{C-E}=1.379, p=.168; Z_{E-H} =1.687, p=.091; Z_{H-C} =-0.474, p=.646). The pretest scores are summarized through descriptive statistics in Table 7.28.

It can be seen from Table 7.28 that the mean scores do differ from section to section. In order to determine whether these differences were significant, or to determine whether the differences could be strictly due to chance, an ANOVA was used. Prior to performing an ANOVA, however, the data were investigated to determine if they fit the assumptions and requirements of the analysis.

The first assumption is that the data are normal. To test this assumption, the standardized residuals (the standardized difference between the observed value and the predicted value based on the general linear model) were plotted in histogram form. This histogram is presented in Figure 7.1. While the data are not strictly normal (the curved line is a normal curve and is given for reference purposes) recall that the ANOVA is robust against this assumption as long as the data are unimodal and somewhat symmetric. The residuals in Figure 7.1 fit this description and so the normal assumption is loosely supported.

Table 7.28 Descriptive Statistics for Pretest Scores, Fall 2002

	Detailed Section	Statistics		Statistic
Pretest Scores	Section C, Fall 2002, Content	N		39
		Mean		29.8205
		95% Confidence Interval for Mean	Lower Bound	27.5424
			Upper Bound	32.0986
		Median		30.5000
		Variance		49.388
		Std. Deviation		7.0277
		Minimum		12.50
		Maximum		39.50
		Range		27.00
	Section E, Fall 2002, Tool-type	N		38
		Mean		27.6184
		95% Confidence Interval for Mean	Lower Bound	25.5181
			Upper Bound	29.7187
		Median		27.7500
		Variance		40.830
		Std. Deviation		6.3899
		Minimum		13.50
		Maximum		40.00
		Range		26.50
	Section H, Fall 2002, Comparison	N		65
		Mean		26.9077
		95% Confidence Interval for Mean	Lower Bound	25.3572
			Upper Bound	28.4582
		Median		27.5000
		Variance		39.155
		Std. Deviation		6.2574
		Minimum		10.50
		Maximum		40.00
		Range		29.50

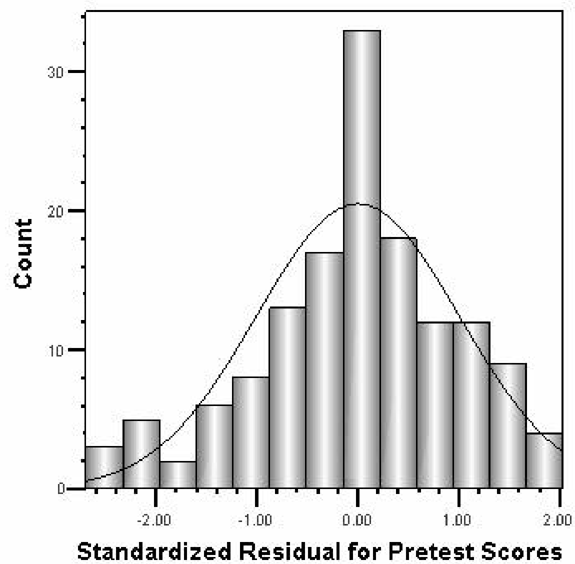


Figure 7.1 Histogram of Standardized Residuals for Pretest Scores, Fall 2002

The second requirement that must be met in order to properly use and interpret the ANOVA is the equality of variances. To test this assumption a Levene test was conducted to compare the pretest score variances from the three sections that participated in the fall of 2002. There was no significant differences between the three variances ($F(2,135)=0.225$, $p=.913$) and so the assumption of equal variances cannot be rejected.

With these two assumptions met, an ANOVA was performed on the pretest scores for the three sections from Fall 2002. The results of the ANOVA are summarized in Table 7.29. As can be seen in the table, the significance level (under the column labeled *Sig.*) is greater than .05, the specified level at which the null hypothesis would be rejected. Thus, the null hypothesis, which for an ANOVA is that there is no difference between the groups, cannot be rejected. The results of the test can be summarized by saying that there are no significant differences between the scores on the pretests from the three different sections that participated in Fall 2002.

Table 7.29 ANOVA of Pretest Scores by Section, Fall 2002

Pretest Scores

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	230.429	2	115.214	2.799	.064
Within Groups	5557.566	135	41.167		
Total	5787.995	137			

For further assurance, an ANOVA was performed also on the time that students took to complete the pretest. Descriptives for the time variable per fall sections are presented in Table 7.30. Note that there are some extreme values in Table 7.30, values as low as 12 minutes and as high as 107 minutes while the means were about 40 minutes (as expected). Such extreme values, called outliers, were removed from the analysis to prevent any bias from students who were not taking the assessment seriously or were trying too hard, to an extent that it was not natural or valid. Four respondents data were removed from the analysis, both for time and for score. The ANOVA results summarized in Table 7.29 do not include these outlying data.

To compare pretest times, ANOVA was once again used once the assumptions had been shown to be valid. Figure 7.2 presents the standardized residuals for the fall data in histogram form. Once again, though the data are not strictly normal, they are unimodal and somewhat symmetric. Furthermore, a Levene’s test was conducted which revealed that there were no significant differences between the variances for the three groups in terms of pretest completion times (F(2,135)=2.699, p=.071). With the ANOVA assumptions satisfied, the analysis was completed as summarized in Table 7.31. The test revealed no significant differences between completion times for the three sections (F(2,135)=.704, p=.496); further evidence that the sections were on equal ground in terms of prior knowledge upon beginning the intervention.

Table 7.30 Descriptive Statistics for Time to Complete Pretest, Fall 2002

	Detailed Section			Statistic
Time to Complete Pretest	Section C, Fall 2002, Content	N		39
		Mean		43.1795
		95% Confidence Interval for Mean	Lower Bound	37.8556
			Upper Bound	48.5034
		Median		45.0000
		Variance		269.730
		Std. Deviation		16.4235
		Minimum		24.00
		Maximum		107.00
		Range		83.00
	Section E, Fall 2002, Tool-type	N		38
		Mean		43.1842
		95% Confidence Interval for Mean	Lower Bound	39.4468
			Upper Bound	46.9216
		Median		42.5000
		Variance		129.289
		Std. Deviation		11.3706
		Minimum		12.00
		Maximum		75.00
		Range		63.00
	Section H, Fall 2002, Comparison	N		64
		Mean		40.3906
		95% Confidence Interval for Mean	Lower Bound	36.4797
			Upper Bound	44.3015
		Median		37.5000
		Variance		245.131
		Std. Deviation		15.6567
		Minimum		19.00
		Maximum		90.00
		Range		71.00

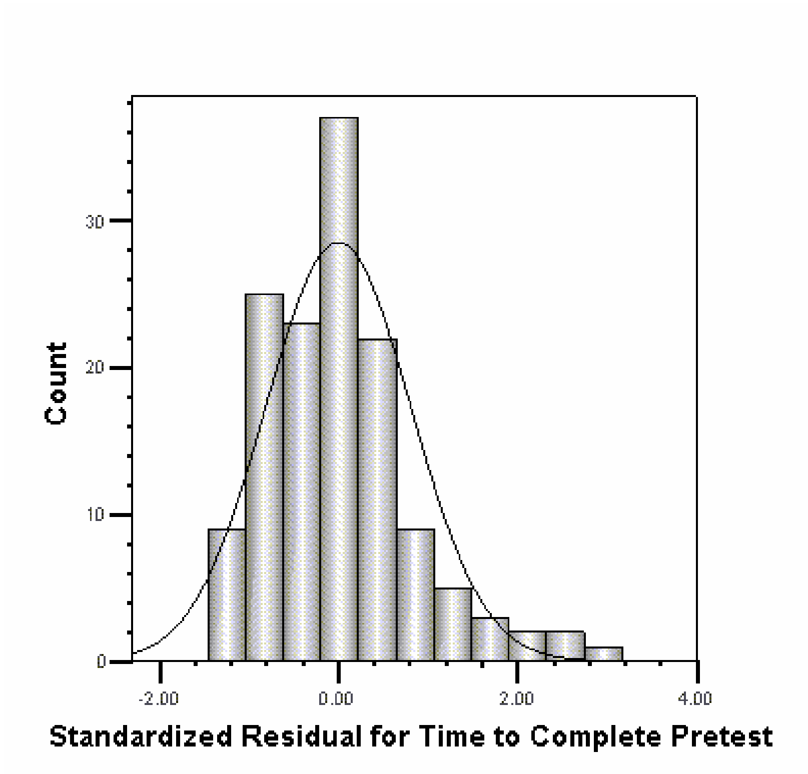


Figure 7.2 Histogram of Standardized Residuals for Time to Complete Pretest, Fall 2002

Table 7.31 ANOVA of Pretest Completion Times by Section, Fall 2002

Time to Complete Pretest

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	258.940	2	129.470	.704	.496
Within Groups	24832.886	135	183.947		
Total	25091.826	137			

7.4.2 Spring 2003 Pretest Data

One hundred and sixteen students from the three participating sections in the spring of 2003 completed the pretest. Twenty-nine of these were from Section B, 44 were from section E, and 43 were from Section G yielding respective return rates of 100, 100, and 96 percents. Two difference of proportions tests were conducted (there is no reason to test the difference between identical proportions) both of which revealed no significant difference in return rates (Z_{E-G}=1.414, p=.159; Z_{G-B} =-1.151, p=.250). Table 7.32 summarizes the pretest scores with descriptive statistics.

Again, an ANOVA was intended to be used to determine if there were any significant differences between pretest scores for the different groups. Again, the assumptions of the ANOVA were tested before employing the method itself. Figure 7.3 is a histogram of the standardized residuals for the pretest score from the spring sections. While the distribution is unimodal, it stretches even the rule of symmetry. Rather than relying on the robustness of the ANOVA in this case, the nonparametric equivalent, the Kruskal-Wallis test, was employed.

The results of this test are summarized in Table 7.33. Note as with other nonparametric tests, the Kruskal-Wallis test is based upon the rank of the scores rather than the scores themselves. These mean ranks are presented in Table 7.33, not the mean scores. The test revealed no significant difference between the pretest scores of the three sections that participated in the spring of 2003 ($?^2(2)$=2.186, p=.335). As such, the sections can be assumed to be equivalent in terms of prior knowledge.

Table 7.32 Descriptive Statistics for Pretest Scores, Spring 2003

	Detailed Section			Statistic
Pretest Scores	Section B, Spring 2003, Content	N		29
		Mean		30.0517
		95% Confidence Interval for Mean	Lower Bound	27.6001
			Upper Bound	32.5034
		Median		30.5000
		Variance		41.542
		Std. Deviation		6.4453
		Minimum		13.50
		Maximum		38.00
		Range		24.50
	Section E, Spring 2003, Tool-type	N		44
		Mean		31.4205
		95% Confidence Interval for Mean	Lower Bound	29.7669
			Upper Bound	33.0740
		Median		33.2500
		Variance		29.581
		Std. Deviation		5.4388
		Minimum		21.50
		Maximum		38.50
		Range		17.00
	Section G, Spring 2003, Comparison	N		43
		Mean		29.7791
		95% Confidence Interval for Mean	Lower Bound	27.6130
			Upper Bound	31.9452
		Median		31.5000
		Variance		49.539
		Std. Deviation		7.0384
		Minimum		13.50
		Maximum		38.50
		Range		25.00

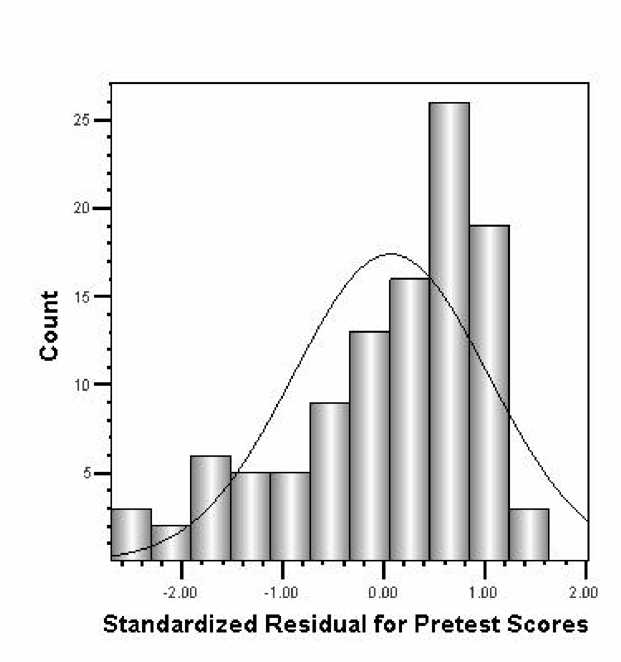


Figure 7.3 Histogram of Standardized Residuals for Pretest Scores, Spring 2003

Table 7.32 Kruskal-Wallis Test for Pretest Scores by Section, Spring 2003

Ranks

	Detailed Section	N	Mean Rank
Pretest Scores	Section B, Spring 2003, Content	28	49.75
	Section E, Spring 2003, Tool-type	39	59.79
	Section G, Spring 2003, Comparison	40	51.33
	Total	107	

Test Statistics[a]

	Pretest Scores
Chi-Square	2.186
df	2
Asymp. Sig.	.335

a. Kruskal Wallis Test

Further evidence of equivalency was sought by comparing pretest completion times. Table 7.33 summarizes the pretest completion time data. As with the fall data, there were extremes in the times, values as high as 120 minutes and as low as 15 minutes. As explained earlier, data with outlying pretest completion times, a total of nine participants, were excluded from the analyses; the Kruskal-Wallis test mentioned in the previous paragraph did not include scores with accompanying extremes in pretest completion times. In an effort to compare the pretest completion times across the three sections, the ANOVA assumptions were again tested. The histogram in Figure 7.4 reveals that the data are not necessarily normally distributed but the distribution is unimodal. Also, with the exception of the high-end outliers, the distribution is somewhat symmetric. In this case, the robustness of the ANOVA was relied upon. A Levene test revealed that there were no significant differences between the variances ($F(2,104)=1.847$, $p=.163$). With the assumptions satisfied, an ANOVA was conducted, the results of which are summarized in Table 7.34. The ANOVA revealed no significant differences between the groups in terms of pretest completion times ($F(2,104)=1.786$, $p=.173$). The results of this test provide further evidence that the groups were equivalent in terms of prior knowledge.

Table 7.33 Descriptive Statistics for Pretest Completion Times, Spring 2003

	Detailed Section			Statistic
Time to Complete Pretest	Section B, Spring 2003, Content	N		28
		Mean		44.7500
		95% Confidence Interval for Mean	Lower Bound	38.8896
			Upper Bound	50.6104
		Median		45.0000
		Variance		228.417
		Std. Deviation		15.1135
		Minimum		25.00
		Maximum		90.00
		Range		65.00
	Section E, Spring 2003, Tool-type	N		42
		Mean		53.3810
		95% Confidence Interval for Mean	Lower Bound	46.2599
			Upper Bound	60.5020
		Median		46.0000
		Variance		522.193
		Std. Deviation		22.8515
		Minimum		15.00
		Maximum		120.00
		Range		105.00
	Section G, Spring 2003, Comparison	N		41
		Mean		44.6341
		95% Confidence Interval for Mean	Lower Bound	39.6117
			Upper Bound	49.6566
		Median		43.0000
		Variance		253.188
		Std. Deviation		15.9119
		Minimum		15.00
		Maximum		85.00
		Range		70.00

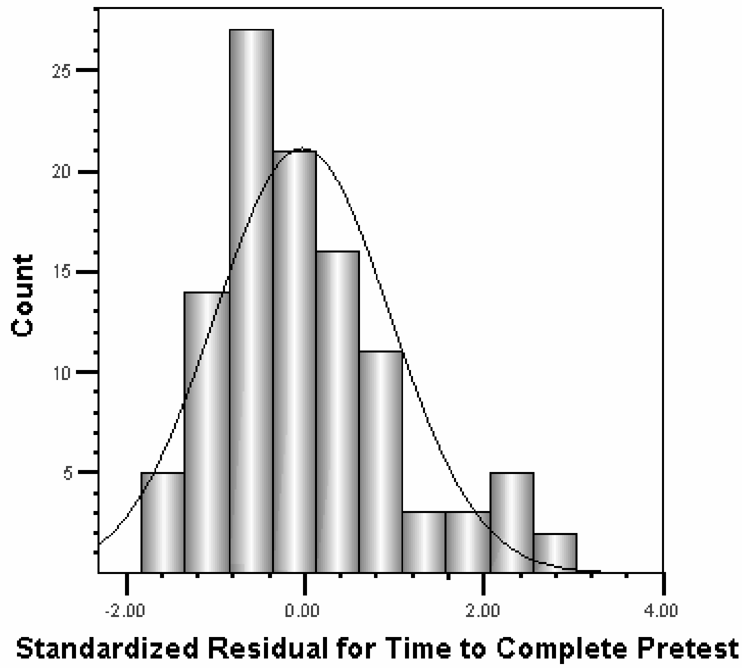


Figure 7.4 Histogram of Standardized Residuals for Pretest Completion Times

Table 7.34 ANOVA for Pretest Completion Times by Section, Spring 2003

Time to Complete Pretest

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	963.998	2	481.999	1.786	.173
Within Groups	28065.292	104	269.859		
Total	29029.290	106			

7.4.3 Combined Pretest Data

As a final comparison, the data were combined into groups based on their respective treatment conditions and again tested for differences. Again, the content-type group include Section C from fall and B from spring, the tool-type group was made up of Sections E from both semesters, and the content-type group was made up of Section H from fall and Section G from spring. The pretest scores for all students are summarized in Table 7.35 below; these data already exclude any completion time outliers that were previously identified as pointed out previously.

An ANOVA was employed to test for any significant differences between the pretest scores for the three treatment conditions. Of course, the ANOVA assumptions were tested first. Figure 5 shows the distribution of the pretest score residuals for all six sections. The distribution is unimodal and reasonably symmetric, thus satisfying the loose requirement of normality. Furthermore, a Levene test revealed that there were no significant differences between the variances for the three treatment conditions ($F(2,242)=2.486$, $p=.085$). With these assumptions in satisfied, the ANOVA was conducted. The analysis, summarized in Table 7.36, revealed that there were no significant differences between the three groups in terms of pretest scores.

A similar analysis was conducted for the pretest completion times for all the sections combined into treatment groups. The pretest completion times broken down by treatment condition are presented in Table 7.37. In order to use ANOVA, the assumptions were first tested. Figure 7.6 shows the distribution of the standardized residuals. Though not normal, they can be assumed to be unimodal and somewhat symmetric. A Levene test revealed no significant difference between the variances ($F(2,242=0.578$, $p=.562$). The ANOVA, which is summarized in Table 7.38, was then conducted which yielded marginally significant differences ($F(2,242)=3.041$, $p=.050$). Recall that in order for the null hypothesis to be rejected, the significance level must be less than the specified rejection value. In this case, the rejection value is .05, for which the p-value above is not less than but equal to, and thus the null is not rejected. Furthermore, post-hoc comparisons, using both the Bonferroni and the Tukey methods, found no statistically significant pairwise differences. Based on this, the assumption that there are no differences between the groups in terms of pretest completion times is still valid.

Table 7.35 Descriptive Statistics for Pretest Scores by Treatment Condition, All Sections

	Treatment Condition			Statistic
Pretest Scores	Content Software	N		66
		Mean		30.1591
		95% Confidence Interval for Mean	Lower Bound	28.5310
			Upper Bound	31.7872
		Median		30.5000
		Variance		43.863
		Std. Deviation		6.6229
		Minimum		12.50
		Maximum		39.50
		Range		27.00
	Tool-Type Software	N		76
		Mean		30.1118
		95% Confidence Interval for Mean	Lower Bound	28.7205
			Upper Bound	31.5031
		Median		29.5000
		Variance		37.071
		Std. Deviation		6.0886
		Minimum		13.50
		Maximum		40.00
		Range		26.50
	Comparison Group	N		103
		Mean		28.2573
		95% Confidence Interval for Mean	Lower Bound	26.9443
			Upper Bound	29.5702
		Median		29.0000
		Variance		45.132
		Std. Deviation		6.7180
		Minimum		10.50
		Maximum		40.00
		Range		29.50

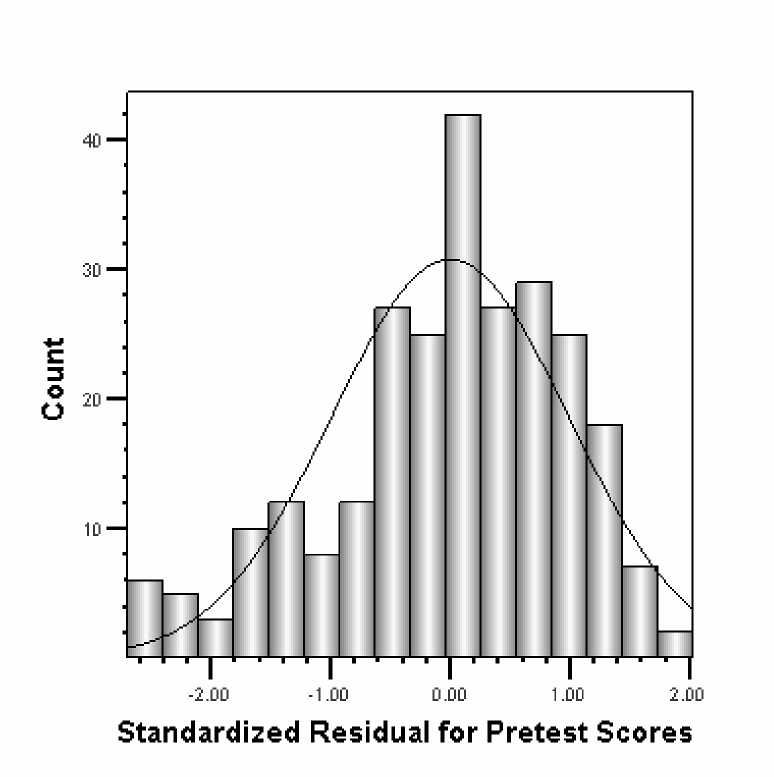


Figure 7.5 Histogram of Standardized Residuals for Pretest Scores, All Sections

Table 7.36 ANOVA of Pretest Scores by Treatment Condition, All Sections

Pretest Scores

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	210.295	2	105.148	2.486	.085
Within Groups	10234.811	242	42.293		
Total	10445.106	244			

Table 7.37 Descriptive Statistics for Pretest Completion Times by Treatment Condition, All Sections

	Treatment Condition			Statistic
Time to Complete Pretest	Content Software	N		66
		Mean		42.8788
		95% Confidence Interval for Mean	Lower Bound	39.4824
			Upper Bound	46.2751
		Median		45.0000
		Variance		190.877
		Std. Deviation		13.8158
		Minimum		24.00
		Maximum		90.00
		Range		66.00
	Tool-Type Software	N		76
		Mean		47.7763
		95% Confidence Interval for Mean	Lower Bound	44.2888
			Upper Bound	51.2638
		Median		44.0000
		Variance		232.923
		Std. Deviation		15.2618
		Minimum		22.00
		Maximum		90.00
		Range		68.00
	Comparison Group	N		103
		Mean		42.5340
		95% Confidence Interval for Mean	Lower Bound	39.4906
			Upper Bound	45.5774
		Median		40.0000
		Variance		242.487
		Std. Deviation		15.5720
		Minimum		20.00
		Maximum		90.00
		Range		70.00

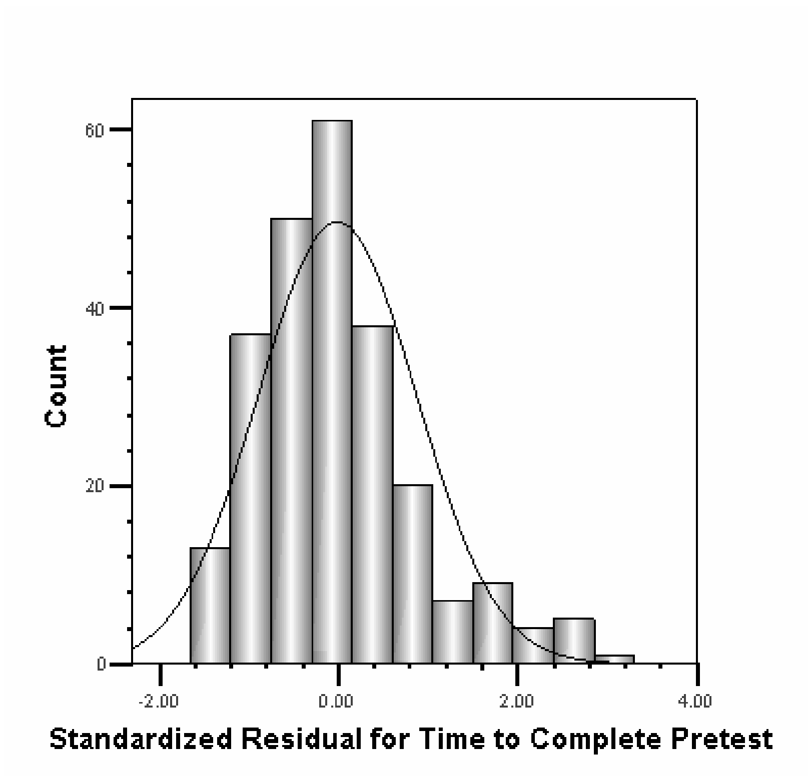


Figure 7.6 Histogram of Standardized Residuals for Pretest Completion Times, All Sections

Table 7.38 ANOVA for Pretest Completion Times by Treatment Condition, All Sections

Time to Complete Pretest

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1372.451	2	686.226	3.041	.050
Within Groups	54609.859	242	225.661		
Total	55982.310	244			

7.4.5 Summary of Pretest Data

The intended purpose in the administration of the pretest was to assess the students prior to the intervention to assure that none of the groups had a significant advantage over any of the other groups in terms of prior knowledge. The results of the various analyses supported the hypothesis that all the groups were similar. The groups within each semester were compared for pretest scores and pretest completion times and no significant differences were found in any of the tests. Groups were then combined based upon their treatment condition and compared across semesters with similar results. As such, it can be concluded that each of the groups entered the truss portion of the course with a similar level of knowledge required to successfully learn trusses and truss analysis.

7.5 Posttest Data

The posttest was administered during class in each of the six sections involved. The students completed the posttest in the lecture period immediately following the special session of the intervention. Students were allowed 40 minutes to complete the test in class. They were informed in advance that their participation grade, five percent of their final grade, would depend on their completing this assignment in class. As with the pretest, the posttests were all graded by the researcher to ensure that there were no differences in scores as a result of grading. There were 55 total points possible on the posttest.

The final version of the posttest, refined as described in Chapter 5 and presented in Appendix C, included five sets of questions. The first set, Problem 1, were factual questions, testing students ability to answer specific questions about trusses such as "True or False: All members of a truss are assumed to be connected by smooth pins." Two question sets were quantitative analysis questions that asked students to solve numerically for the internal forces in two different trusses and state whether the forces were in tension or compression. While students could use either the method of sections or the method of joints, or a combination of the two, to solve for either of the trusses, Problem 2 lent itself more to solving via the method of joints and the other, Problem 4, lent itself to solving via the method of sections.

Two other sets of questions required students to qualitatively analyze trusses. These questions, as mentioned in detail in Chapter 5, were different then the types of questions that students usually encounter when studying trusses. These questions required students to think about the truss and the interaction of its parts without actually solving it numerically. The students were assigned no homework of this type and no questions of this type were covered in class. Interviews with statics instructors, however, revealed that was a desirable skill or ability. Because of this, students did see this type of question on the assignment that was completed during the special session to determine if the software would help students in this area. As such, it was also necessary to include this type of question on the posttest to see if an effect of this kind actually occurred.

Throughout the remainder of this chapter, these question types will be referred to as factual questions (Problem 1), quantitative questions (Problems 2 and 4), and qualitative questions (Problems 3 and 5) respectively. The analyses and results of the posttest data will be presented in a manner similar to that of the pretest data. Results from the Fall 2002 sections will be presented first, followed by data from the following spring, and combined data will be presented thereafter.

7.5.1 Fall Posttest Data

One hundred and thirty-six students from the fall sections of CEE 2020 completed the posttest; 37 of these were from Section C, 40 were from Section E, and 59 were from Section H. This translates into respective completion rates of 82, 95, and 83 percents. Three difference of proportions tests revealed no significant differences between any of the return rates ($Z_{C\text{-}E}$=-1.902, p=.057; $Z_{E\text{-}H}$ =1.893, p=.059; $Z_{H\text{-}C}$ =0.212, p=.834). The pretest scores are summarized via descriptive statistics in Table 7.39.

7.5.1.1 Total Scores on Posttest

It can be seen from the table that Section C, the content-type software group, has a mean that is somewhat higher than the other two sections. Section C also has a much higher variance than the other two sections as well. It is impossible to tell from these descriptive data whether there are any actual differences between the sections. In order to make this determination, a number of inferential techniques were used. In order to implement the use of parametric statistics, the assumption of normality must be met. Figure 7.7 shows the histogram of the standardized residuals for the posttest scores. Again, the data are clearly not normal, but it does have a unimodal, curvy shape which is sufficient, especially with relatively large sample populations as is the case with this experiment.

As opposed to the pretest however, where the hypothesis was that there were no differences between the scores for the different sections, different posttest scores have been hypothesized. More specifically, it was hypothesized that the experimental sections would perform better than the comparison group and that the two experimental sections would perform equally well.

An orthogonal set of planned contrasts was designed to test these hypotheses. The two contrasts, with the coefficients shown in Table 7.40, did not support the hypotheses. The first contrast tested whether the average of the mean scores for the experimental groups (Sections C and E) was equal to the mean score for the comparison group. The second contrast tested whether the experimental sections had the same means. In essence, conducting the planned comparisons allowed for the simultaneous testing of both posttest hypotheses.

The first comparison revealed that there was a significant difference between the average of the means for the experimental groups and the mean for the comparison group, with the experimental groups scoring higher ($t(127.897)=2.222$, $p<.05$). Note that in the test above, and in tests that follow, a t-test that does not assume equality of variances was used, which is why the degrees of freedom is not an integer. The second comparison revealed that the content-type software group (Section C) performed significantly better on the posttest ($t(70.055)=2.524$, $p<.05$) than did the tool-type software group (Section E). Based on these results, the hypothesis that both experimental sections would perform equally well on the posttest must be rejected. Action on the first hypothesis, though, required more investigation because it was unclear whether the tool-type group performed significantly different then the comparison group. Had the two experimental groups performed equally well and the combination of the two been higher than the control group, it could be concluded that both groups did better than the control. In this case, however, it is possible that the content-type group mean is so high that it is inflating the average of the two experimental means.

To determine the nature of the relationship between the tool-type group and the control group required another comparison. This comparison is considered a post-hoc comparison because it was conducted only after certain significant relationships had already been determined. A t-test was conducted to compare the tool-type experimental group to the control group; the results of this test revealed that there was no significant difference between the two groups ($t(79.072)=0.339$, $p=.736$).

The results of these tests can be summarized as follows: there was no significant difference between posttest scores for the control and tool-type group, both of which were significantly lower than the content-type group scores.

Table 7.39 Descriptive Statistics for Posttest Scores by Sections of CEE 2020, Fall 2002.

	Section ID			Statistic
Score on Posttest	Section C, Fall 2002	N		37
		Mean		22.973
		95% Confidence Interval for Mean	Lower Bound	19.452
			Upper Bound	26.494
		Median		20.000
		Variance		111.527
		Std. Deviation		10.561
		Minimum		5.5
		Maximum		43.0
		Range		37.5
	Section E, Fall 2002	N		40
		Mean		17.375
		95% Confidence Interval for Mean	Lower Bound	14.583
			Upper Bound	20.167
		Median		15.250
		Variance		76.202
		Std. Deviation		8.729
		Minimum		3.0
		Maximum		37.0
		Range		34.0
	Section H, Fall 2002	N		59
		Mean		16.788
		95% Confidence Interval for Mean	Lower Bound	14.696
			Upper Bound	18.880
		Median		15.000
		Variance		64.459
		Std. Deviation		8.029
		Minimum		3.0
		Maximum		37.0
		Range		34.0

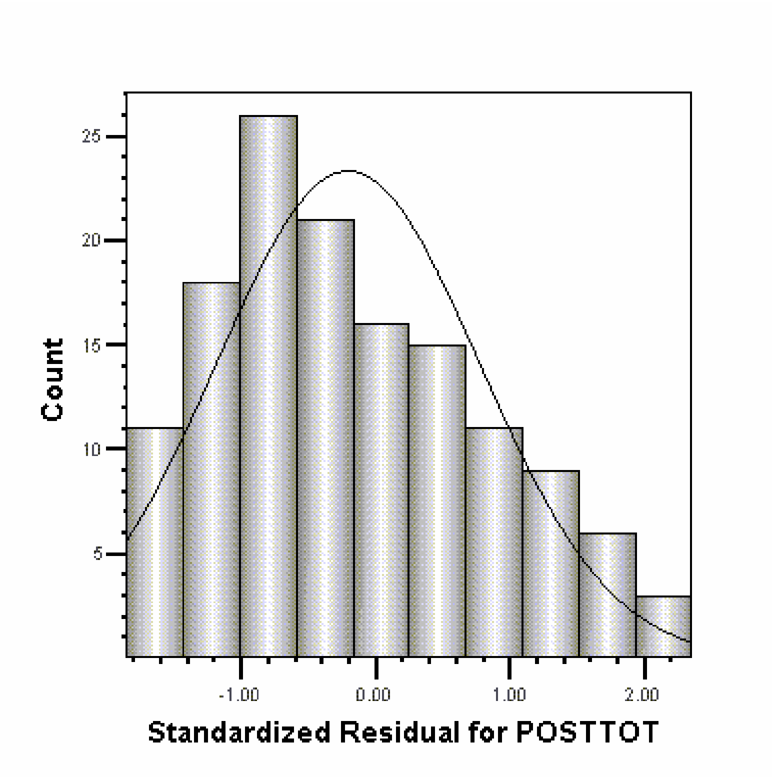


Figure 7.7 Histogram of Standardized Residuals for Posttest Scores, Fall 2002

Table 7.40 Contrast Coefficients for Posttest Hypothesis Testing

	Section ID		
Contrast	Section C, Fall 2002	Section E, Fall 2002	Section H, Fall 2002
1	1	1	-2
2	1	-1	0

During the research design phase of this research, no specific hypotheses were formed about types of questions that would appear on the posttest. As the research progressed, however, it became clear that this type of data might be illustrative, especially since the statics instructors believed that the software would help students in a particular area: the qualitative analysis of trusses. Because of this, further analysis was conducted on the posttest data, which was broken down by question type.

7.5.1.2 Scores on Individual Question Types

A total of five points were possible on Problem 1 of the posttest with no partial points being awarded by the grader. As such, these data cannot be considered continuous but each measurement would fall into one of six categories: 0, 1, 2, 3, 4, or 5. A chi-square test was therefore used to compare the distributions of scores on factual questions between the fall sections of CEE 2020. The data used in this comparison are presented in Table 7.41; note that scores of 0, 1, and 2 were combined into a single category in order to meet the requirements of the chi-square test. The test revealed no significant differences between the three sections on Problem 1 scores ($\chi^2(6)=11.416$, $p=.076$).

Table 7.41 Posttest Problem 1 Score by Sections of CEE 2020, Fall 2002.

			Score on Problem 1 on Posttest				
			<3.0	3.0	4.0	5.0	Total
Section ID	Section C, Fall 2002	Count	4	9	9	15	37
		% within Section ID	10.8%	24.3%	24.3%	40.5%	100.0%
	Section E, Fall 2002	Count	1	11	19	9	40
		% within Section ID	2.5%	27.5%	47.5%	22.5%	100.0%
	Section H, Fall 2002	Count	7	21	13	18	59
		% within Section ID	11.9%	35.6%	22.0%	30.5%	100.0%
Total		Count	12	41	41	42	136
		% within Section ID	8.8%	30.1%	30.1%	30.9%	100.0%

The scores on the quantitative questions, presented in Table 7.42, do not fit the ANOVA requirements. The data are not normal, nor are the variances homogeneous ($F(2,133)=5.063$, $p<.05$). For this reason, a KW test was conducted which revealed that significant differences did occur ($\chi^2(2)=8.363$, $p<.05$). Post-hoc comparisons between the mean ranks revealed the nature of these differences. Three post-hoc comparisons were made with a Bonferroni correction yielding critical pairwise comparison values of $a=.017$ and $Z_c=2.394$. Only one significant relationship existed, the content-type group (Section C) performed significantly better than the control group (Section H) on the quantitative questions ($Z=2.84$, $p<.017$). The relationship between the tool-type group and the content-type group was not significant ($Z=2.02$, $p=.043$) nor was the relationship between the tool-type group and the control group ($Z=0.66$, $p=.509$). In summary, on the quantitative questions, the content-type group performed better than the control group and the tool-type group performed equally as well as the other two groups.

A similar comparison was made with the combined scores on the two qualitative analysis questions; these scores are presented in Table 7.43. As with the quantitative data, normality could not be assumed and the variances were not equal ($F(2,133)=2.296$, $p=.023$). Thus, another KW test was completed for the qualitative data, revealing no significant differences between any of the groups ($\chi^2(2)=5.672$, $p=.059$). It can therefore be concluded that all groups performed equally well on the qualitative posttest questions.

Table 7.42 Descriptive Statistics for Scores on Quantitative Posttest Questions, Fall 2002.

Quantitative Posttest Questions

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Section C, Fall 2002	37	12.5676	8.0710	1.3269	9.8766	15.2586	.00	24.00
Section E, Fall 2002	40	9.1000	7.8334	1.2386	6.5948	11.6052	.00	23.00
Section H, Fall 2002	59	7.7373	6.3289	.8240	6.0880	9.3866	.00	23.00
Total	136	9.4522	7.5026	.6433	8.1799	10.7245	.00	24.00

Table 7.43 Descriptive Statistics for Scores on Qualitative Posttest Questions, Fall 2002.

Qualitative Posttest Questions

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Section C, Fall 2002	37	6.4595	3.9341	.6468	5.1478	7.7712	.00	14.00
Section E, Fall 2002	40	4.3750	2.7333	.4322	3.5008	5.2492	.00	10.00
Section H, Fall 2002	59	5.3559	3.1881	.4151	4.5251	6.1868	.00	15.00
Total	136	5.3676	3.3573	.2879	4.7983	5.9370	.00	15.00

7.5.1.3 Summary of Fall Posttest Data

There were a few significant differences for the fall posttest data. First, the content-type groups performed better on the overall posttest than either of the other groups. Second, the content-type group performed better on the on the quantitative questions than the control group. These results and their implications will be discussed in greater detail in the following chapter. It is important to note that these results are not valid alone. As mentioned previously, replication is an necessary part of a quasi-experimental design in order to avoid any sampling bias. Thus, similar results from the replicated study are necessary to validate these results.

Recall also that the distributions of students by rank were not equal among the three fall sections. Rather than including rank as a factor in the analysis to overcome this possible bias, replicated results would be relied upon because the distributions of students by rank were equal in the spring sections. If similar results were realized in the spring posttest data, then rank could be ruled out as a reason for the differences above. The posttest data from the replicated study, Spring 2003, is presented in the next section.

7.5.2 Spring Posttest Data

One hundred and seven students from the fall sections of CEE 2020 completed the posttest; 23 of these were from Section B, 42 were from Section E, and 42 were from Section G. This translates into respective completion rates of 79, 91, and 88 percents. Three difference of proportions tests revealed no significant differences between any of the return rates (Z_{B-E}=-1.488, p=.136; Z_{E-G} =0.598, p=.549; Z_{G-B} =0.960, p=.337). The pretest scores are summarized via descriptive statistics in Table 7.44.

7.5.2.1 Total Scores on Posttest

The hypotheses in the replicated study were identical to those in the initial study. The same set of contrasts was used to test these hypotheses once the normality assumption was satisfied. Figure 7.8 is the histogram of the standardized residuals for the posttest scores. This distribution definitely stretches the assumption of normality, but the sample sizes are sufficient that the robustness of the tests against the normality assumption will be relied upon. The t-test that is used to compare the contrasts does not require equality of variances so that assumption was not tested.

The contrast coefficients used to test the hypotheses are identical to those used in Fall 2002 and are presented in Table 7.45. The first contrast revealed no significant difference between the average of the experimental means and the control mean ($t(80.428)=-1.158$, $p=.760$). The second contrast revealed no significant difference between the means of the experimental groups either ($t(39.535)=-2.468$, $p=.347$). Furthermore, based on these results it can be concluded that there is no significant difference between the posttest scores of any of the groups.

Further analysis of the posttests by question types was conducted for the spring data as well.

7.5.2.2 Scores on Individual Question Types

As with the data from the previous semester, the distributions of scores on Problem 1 of the posttest, shown in Table 7.46, were compared with using a chi-square test. The test revealed that there was a significant difference in the distributions ($\chi^2(6)=14.454$, $p<.05$). Post-hoc analyses of the adjusted residuals, with a Bonferroni correction for 12 comparisons ($a=.00417$, $Z_c=2.866$), revealed one significant relationship: significantly more students in Section G (the control group) received a score of 3 on Problem 1 than was expected based on the distributions of scores ($Z=2.878$, $p<.00417$). Neither the reasons for nor the implications of this slightly significant relationship are known or assumed. This results implies that the control group performed significantly worse on the factual questions than did the experimental groups.

Table 7.47 presents the scores on the quantitative questions of the pretest for each of the three spring sections and Table 7.48 presents the qualitative data. Two KW tests were conducted, which revealed that no significant differences occurred between the three sections on the quantitative ($\chi^2(2)=3.721$, $p=.156$) or qualitative questions ($\chi^2(2)=5.678$, $p=.058$).

Table 7.44 Descriptive Statistics for Posttest Scores for Sections of CEE 2020, Spring 2003

	Section ID			Statistic
Score on Posttest	Section B, Spring 2003	N		23
		Mean		21.413
		95% Confidence Interval for Mean	Lower Bound	16.859
			Upper Bound	25.967
		Median		23.000
		Variance		110.924
		Std. Deviation		10.532
		Minimum		5.0
		Maximum		39.0
		Range		34.0
	Section E, Spring 2003	N		42
		Mean		23.881
		95% Confidence Interval for Mean	Lower Bound	21.091
			Upper Bound	26.671
		Median		23.500
		Variance		80.156
		Std. Deviation		8.953
		Minimum		8.0
		Maximum		40.0
		Range		32.0
	Section G, Spring 2003	N		42
		Mean		23.226
		95% Confidence Interval for Mean	Lower Bound	20.459
			Upper Bound	25.993
		Median		23.250
		Variance		78.856
		Std. Deviation		8.880
		Minimum		5.0
		Maximum		39.5
		Range		34.5

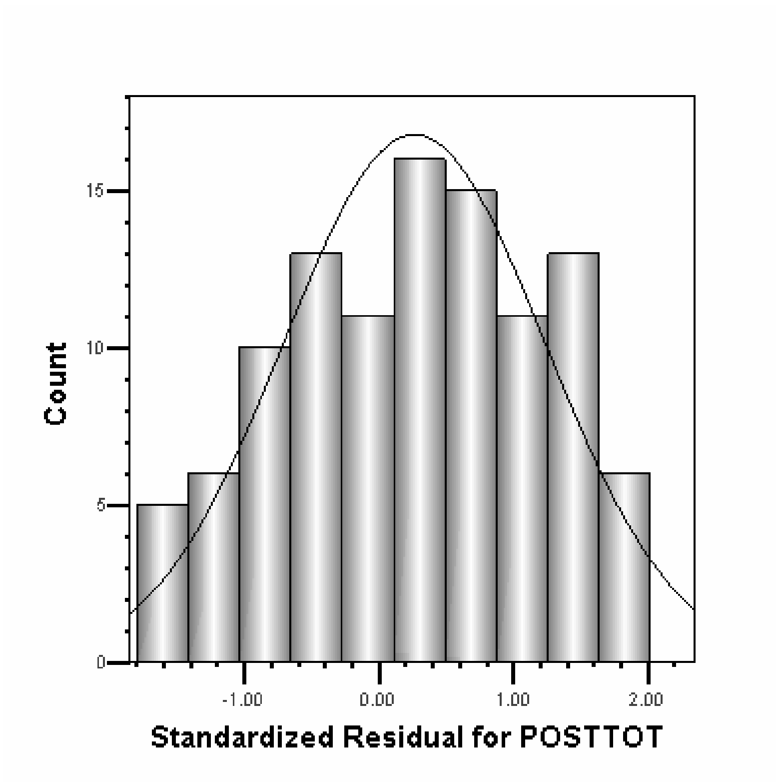


Figure 7.8 Histogram of Standardized Residuals for Posttest Scores, Spring 2003

Table 7.45 Contrast Coefficients to Test Posttest Hypotheses, Spring 2003

Contrast Coefficients

	Section ID		
Contrast	Section B, Spring 2003	Section E, Spring 2003	Section G, Spring 2003
1	1	1	-2
2	1	-1	0

Table 7.46 Posttest Problem 1 Scores by Sections of CEE 2020, Spring 2003

			Score on Problem 1 on Posttest				
			<3.0	3.0	4.0	5.0	Total
Section ID	Section B, Spring 2003	Count	4	1	8	10	23
		% within Section ID	17.4%	4.3%	34.8%	43.5%	100.0%
	Section E, Spring 2003	Count	2	7	16	17	42
		% within Section ID	4.8%	16.7%	38.1%	40.5%	100.0%
	Section G, Spring 2003	Count	7	15	10	10	42
		% within Section ID	16.7%	35.7%	23.8%	23.8%	100.0%
Total		Count	13	23	34	37	107
		% within Section ID	12.1%	21.5%	31.8%	34.6%	100.0%

Table 7.47 Descriptive Statistics for Posttest Quantitative Questions, Spring 2003

Quantitative Posttest Questions

					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Section B, Spring 2003	23	11.9783	7.7232	1.6104	8.6385	15.3180	.00	24.00
Section E, Spring 2003	42	15.6429	6.6821	1.0311	13.5606	17.7251	1.00	24.00
Section G, Spring 2003	42	14.3452	7.0022	1.0805	12.1632	16.5273	.00	23.50
Total	107	14.3458	7.1061	.6870	12.9838	15.7078	.00	24.00

Table 7.48 Descriptive Statistics for Posttest Qualitative Questions, Spring 2003

Qualitative Posttest Questions

					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Section B, Spring 2003	23	5.4348	3.4487	.7191	3.9435	6.9261	.00	12.00
Section E, Spring 2003	42	4.0952	3.3262	.5133	3.0587	5.1318	.00	13.00
Section G, Spring 2003	42	5.4048	3.1083	.4796	4.4362	6.3734	2.00	13.00
Total	107	4.8972	3.3022	.3192	4.2643	5.5301	.00	13.00

7.5.2.3 Summary of Spring Posttest Data

Only one significant relationship was revealed in the posttest data from Spring 2003: the control group performed worse on the factual questions of the posttest than either of the experimental groups. All other relationships revealed no significant difference.

7.5.3 Comparison of Fall and Spring Posttest Results

As mentioned previously, replication was used in this study in order to remove the possibility of a sample bias that might be present due to a lack of random selection or assignment. As a result of this replication, only significant results that are replicated are valid and generalizable. No significant results were realized in both applications of this study.

One reason for the significant difference in overall posttest scores that was realized in the fall but not the spring may be the differing distributions of students by rank that was present in the fall but not the spring. To test this hypothesis, a factorial analysis, or two-way ANOVA, was conducted on the fall pretest scores with section and rank as the factors. The data for this test have already been assumed to be normal in previous tests and a Levene's Test reveled no significant difference between the variances ($F(8,127)=1.736$, $p=.096$), thus satisfying the requirements for this analysis. The data for this analysis are summarized in Table 7.49 and the test is summarized in Table 7.50.

While the results in Table 7.50 reveal no significant difference due to rank ($F(2,127)=1.415$, $p=.247$) or due to an interaction between rank and section ($F(4,127)=1.436$, $p=.226$), the partitioning of the sum of the squares isolated enough of the rank and/or interaction variability out of the section variability such that there was no longer a main effect for section. What this means is that when rank is accounted for, there are no significant differences between the posttest scores for the three groups ($F(2,127)=2.396$, $p=.095$). These results are consistent with results from the replicated study in the spring, which revealed no significant differences in the posttest scores among the groups and no significant differences in the population distributions of students by rank.

Possible reasons for the differences in question types were not investigated. Because none of these relationships was replicated these results were not considered valid. As such, it was concluded that, overall, there were no significant differences between the scores for the groups on any of the three question types.

Table 7.49 Descriptive Statistics for Posttest Scores by Section and Rank, Fall 2002

Dependent Variable: Score on Posttest

Section ID	Rank	Mean	Std. Deviation	N
Section C, Fall 2002	Sophomore	21.333	10.959	12
	Junior	24.786	10.803	21
	Senior	18.375	7.609	4
	Total	22.973	10.561	37
Section E, Fall 2002	Sophomore	21.038	10.142	13
	Junior	15.891	7.934	23
	Senior	14.000	5.228	4
	Total	17.375	8.729	40
Section H, Fall 2002	Sophomore	15.417	9.068	12
	Junior	19.639	8.556	18
	Senior	15.586	7.018	29
	Total	16.788	8.029	59
Total	Sophomore	19.311	10.179	37
	Junior	19.992	9.781	62
	Senior	15.716	6.822	37
	Total	18.643	9.304	136

Table 7.50 Summary of Factorial Analysis of Pretest Scores by Section and Rank, Fall 2002

Dependent Variable: Score on Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1628.251[a]	8	203.531	2.570	.012
Intercept	29822.666	1	29822.666	376.556	.000
SECT_ID	379.576	2	189.788	2.396	.095
RANK	224.143	2	112.072	1.415	.247
SECT_ID * RANK	454.993	4	113.748	1.436	.226
Error	10058.203	127	79.198		
Total	58956.750	136			
Corrected Total	11686.454	135			

a. R Squared = .139 (Adjusted R Squared = .085)

7.5.4 Combined Posttest Data

As a final analysis of the posttest data, sections that received the same treatment in the intervention were lumped together to analyze all the data at once. Sections C from fall and B from spring were combined into the content-type group, sections E from each semester were combined into the tool-type group, and sections H from fall and G from spring were combined into the control group. The combined posttest data are presented in Table 7.51.

The same planned contrasts that were used to test the main hypotheses in the individual semesters as described above were used to compare the combined posttest data. There was no significant difference between the average of the means for the experimental groups and the control group mean ($t(215.119)=1.685$, $p=.093$). There was also no significant difference between the mean score for the content-type group and the mean score for the tool-type group ($t(118.658)=0.978$, $p=.330$). It can be concluded from these results that there are no significant differences between pretest scores of any of the three treatment groups.

Comparison for scores on individual question types were not conducted because the combined data were not meant to be a primary source of information. The combined data cannot be replicated and therefore cannot be validated against possible sampling biases. The analysis of the total scores by combined sections is included here simply for illustrative purposes and should not be interpreted as being more accurate or appropriate than the replicated results as originally designed.

Table 7.51 Descriptive Statistics for Posttest Scores by Treatment Condition

Score on Posttest

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Content-type Software Group	60	22.375	10.488	1.354	19.666	25.084	5.0	43.0
Tool-type Software Group	82	20.707	9.379	1.036	18.646	22.768	3.0	40.0
Comparison Group	101	19.465	8.938	.889	17.701	21.230	3.0	39.5
Total	243	20.603	9.518	.611	19.400	21.806	3.0	43.0

7.5.5 Summary of Posttest Data

The posttest was designed to be a standard measure of truss learning for the purposes of this experiment. The posttest was completed in class immediately following the truss portion of the course and the intervention. The posttest scores were compared here across different sections that received different experimental treatments. No significant difference in total posttest scores between groups occurred in either implementation of the study. The total posttest scores were also broken down by question type and compared across groups, these comparisons also revealed no significant differences. It can be concluded that treatment which the groups received during the intervention did not have an effect on learning, nor did it have an effect on what type of problems students would be able to solve. All the groups learned equal amounts and all were equally prepared to answer each of they three types of questions.

7.6 Exam Questions

Midterm exam questions were administered in each of the semesters as described in the previous chapter. The exams given to the different semesters were quite different, because spring students would have copies of the previous fall exams. As such, no comparison should be made between the scores for the different semesters. As with all the assessment instruments, the exam questions were graded by the researcher with great care taken to ensure that each exam was graded according to the same standard. The results from these exams are in presented in Tables 7.52 and 7.53 for fall and spring respectively.

The distributions of the data, see Figure 7.9, are not normal and cannot even be assumed to smooth and humpy-shaped. The distributions, however, are quite similar to the distribution of exam question results from the formative study (see Figure 5.5). Statics instructors who were questioned about this phenomenon revealed that these bimodal distributions are not unusual on exam questions involving quantitative analysis of trusses. Such distributions do, however, prevent the use of parametric statistical methods.

Table 7.52 Descriptive Statistics for Exam Questions by Section of CEE 2020, Fall 2002

Exam Questions

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Section C, Fall 2002	40	19.6500	6.6624	1.0534	17.5193	21.7807	6.00	25.00
Section E, Fall 2002	40	19.6250	6.5111	1.0295	17.5427	21.7073	4.00	25.00
Section H, Fall 2002	67	19.5373	6.8497	.8368	17.8665	21.2081	2.00	25.00
Total	147	19.5918	6.6629	.5495	18.5057	20.6779	2.00	25.00

Table 7.52 Descriptive Statistics for Exam Questions by Section of CEE 2020, Spring 2003

Exam Questions

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Section B, Spring 2003	29	18.4828	4.8743	.9051	16.6287	20.3368	8.00	25.00
Section E, Spring 2003	43	20.2093	4.6882	.7149	18.7665	21.6521	5.00	25.00
Section G, Spring 2003	42	19.9762	5.0388	.7775	18.4060	21.5464	8.00	25.00
Total	114	19.6842	4.8760	.4567	18.7794	20.5890	5.00	25.00

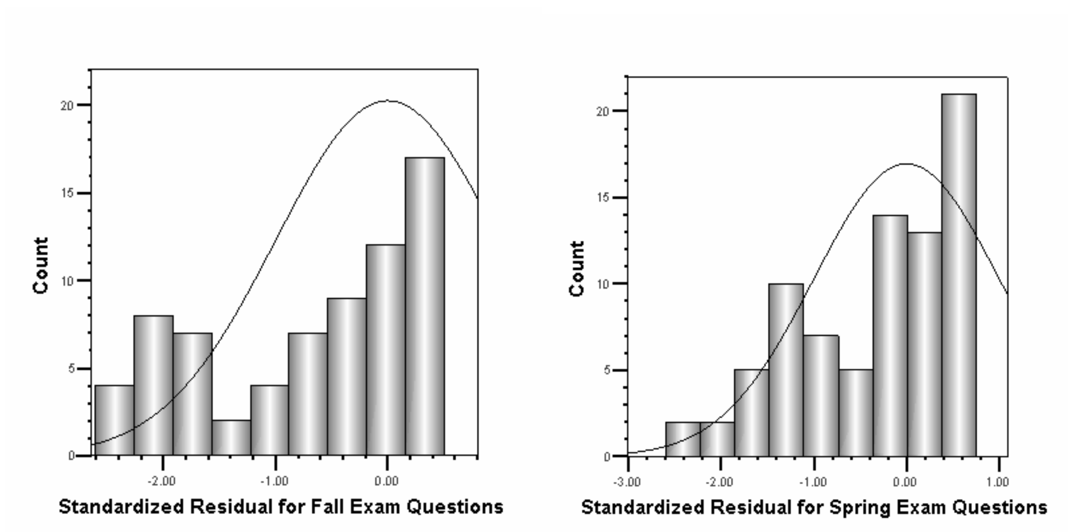


Figure 7.9 Histograms of Residuals for Exam Questions

It can be seen from Table 7.52 that the scores varied little between the different sections; both the means and the standard deviations are almost identical. The nonparametric KW test was used to compare the scores and no significant difference was revealed (?^{2}(2)=0.014, p=.993). The spring results varied slightly more than the fall exam scores but the KW test revealed that there was still no significant difference between the scores for the different sections (?^{2}(2)=3.056, p=.217).

It can be concluded from these results that there was no significant difference between the scores on the exam questions for the different sections in both the fall and spring semesters. Furthermore, it can be concluded that the experimental treatments did not affect, either positively or negatively, the students' ability to quantitatively analyze trusses in an exam setting.

7.7 Retention Data

The retention test was administered approximately ten weeks after the posttest. Students took the test home to complete it; They were allowed one week to complete the exam but they were explicitly instructed to spend no more than forty minutes on the assignment. With the exception of a demographic questionnaire, the retention test was identical to the posttest. There were 55 total points available on the retention test and the exam was graded by the researcher according to the same standards that were used to grade all the posttests, retention tests, and long-term retention tests. This section will present the results of the pretest data by semesters looking at total scores, differential scores, and scores on different question types.

7.7.1 Fall Retention Data

One hundred and thirty-seven students completed the retention test: 37 of these were from Section C, 37 were from Section E, and 63 were from Section H. These numbers translate into respective return rates of 90, 97, and 90 percents respectively. Three difference of proportions tests revealed no significant differences between any of the return rates (Z_{C-E}=-1.299, p=.194; Z_{E-H} =1.396, p=.160; Z_{H-C} =-0.041, p=.968).

7.7.1.1 Total Scores on Retention Test

The retention test scores are summarized in Table 7.53. Ideally, the same contrasts that were used with the posttest would have been used on the retention test. The retention data from this semester, however, cannot be assumed to be normal or even smooth and unimodal (see Figure 7.10). Without the normality assumption satisfied, the t-test used to analyze the planned contrasts is not valid and should not be used. As a result, a KW test was used to compare the total retention scores of all three sections at once. The results of the KW test revealed that there were no significant difference in retention test scores for any of the sections ($\chi^2(2)=5.321$, $p=.070$). Thus, it can be concluded that all the sections performed equally well on the retention test.

Table 7.53 Descriptive Statistics for Retention Test Scores by Section, Fall 2002

Score on Retention Test

					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Section C, Fall 2002	37	24.257	10.463	1.720	20.768	27.745	3.0	43.0
Section E, Fall 2002	37	21.135	9.724	1.599	17.893	24.377	6.0	46.5
Section H, Fall 2002	63	19.270	11.966	1.508	16.256	22.283	1.0	45.0
Total	137	21.120	11.115	.950	19.243	22.998	1.0	46.5

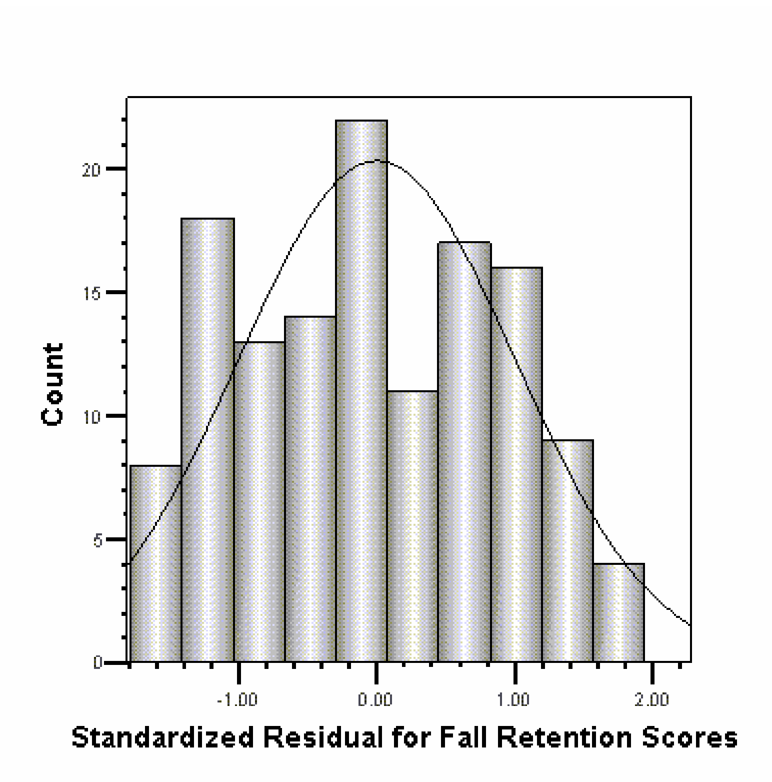


Figure 7.10 Histogram of Standardized Residuals for Retention Test Scores, Fall 2002

7.7.1.2 Scores on Individual Question Types

Further analysis on the fall retention data was completed by breaking down the total scores by individual question types. As with the posttest data, a chi-square test was used in the analysis of the factual questions and KW tests were used for the quantitative and qualitative questions.

The results of Problem 1 from the retention test, the factual questions, are presented in Table 7.54, recall that there were five points possible on Problem 1 with no partial credit awarded. All scores below 3.0 were lumped into a single category to allow for appropriate use of the chi-square test. The chi-square test revealed no significant difference between the scores for the three sections on Problem 1 of the pretest ($\chi^2(6)=5.917$, p=.433).

Table 7.54 Scores on Problem 1 of Retention Test by Section, Fall 2002

			Score on Problem 1 on Retention Test				
			<3.0	3.0	4.0	5.0	Total
Section ID	Section C, Fall 2002	Count	3	12	16	6	37
		% within Section ID	8.1%	32.4%	43.2%	16.2%	100.0%
	Section E, Fall 2002	Count	5	6	18	8	37
		% within Section ID	13.5%	16.2%	48.6%	21.6%	100.0%
	Section H, Fall 2002	Count	10	21	20	12	63
		% within Section ID	15.9%	33.3%	31.7%	19.0%	100.0%
Total		Count	18	39	54	26	137
		% within Section ID	13.1%	28.5%	39.4%	19.0%	100.0%

The results of the qualitative questions are summarized in Table 7.55. A KW test revealed no significant differences between the qualitative scores for the three different fall sections ($\chi^2(2)=5.757$, p=.057). As such, it was concluded that each of the sections performed equally well on the qualitative questions.

Table 7.55 Descriptive Statistics for Qualitative Retention Test Scores, Fall 2002

Score on Qualitative Retention Test Questions

					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Section C, Fall 2002	36	6.7222	3.4774	.5796	5.5457	7.8988	.00	16.00
Section E, Fall 2002	37	5.1351	3.4654	.5697	3.9797	6.2906	.00	19.00
Section H, Fall 2002	63	5.4603	4.6449	.5852	4.2905	6.6301	.00	21.00
Total	136	5.7059	4.0791	.3498	5.0141	6.3976	.00	21.00

The results of the quantitative questions are summarized in Table 7.56. A KW test revealed no significant difference on the quantitative question scores between the three sections involved in the fall study ($\chi^2(6)=3.873$, $p=.144$). As with the previous question types, it can be concluded that the different sections performed equally well on the quantitative questions regardless of their differing experimental treatments.

Table 7.56 Descriptive Statistics for Scores on Quantitative Retention Test Questions, Fall 2002.

Score on Quantitative Retention Test Questions

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Section C, Fall 2002	37	13.7703	8.4423	1.3879	10.9555	16.5851	.00	24.00
Section E, Fall 2002	37	12.2162	7.4996	1.2329	9.7157	14.7167	.00	22.50
Section H, Fall 2002	63	10.3333	8.5180	1.0732	8.1881	12.4786	.00	24.00
Total	137	11.7701	8.3024	.7093	10.3673	13.1728	.00	24.00

7.7.1.3 Differential Scores

As described in previous chapters, one other measure of retention is the differential score or the difference between the score on the retention test and the score on the posttest. Differential scores are intended to show how much knowledge or ability is lost or gained in the interval between the two assessments. The differential scores for retention are summarized in Table 7.57 and a histogram of the standardized residuals is presented in Figure 7.11. It can be seen in Table 7.57 that scores ranged from moderately negative to highly positive (note that with a total possible points of 55 on the retention test, the range of possible differential scores is –55 to 55). The ranges, standard deviations, and means were somewhat similar for each of the three groups.

Table 7.57 Descriptive Statistics for Retention Differential Scores, Fall 2002

Retention Differential Score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Section C, Fall 2002	34	1.9706	8.5324	1.4633	-1.0065	4.9477	-14.00	26.00
Section E, Fall 2002	37	3.2973	8.6910	1.4288	.3996	6.1950	-13.00	21.50
Section H, Fall 2002	55	3.3000	10.4950	1.4151	.4628	6.1372	-18.00	32.50
Total	126	2.9405	9.4290	.8400	1.2780	4.6029	-18.00	32.50

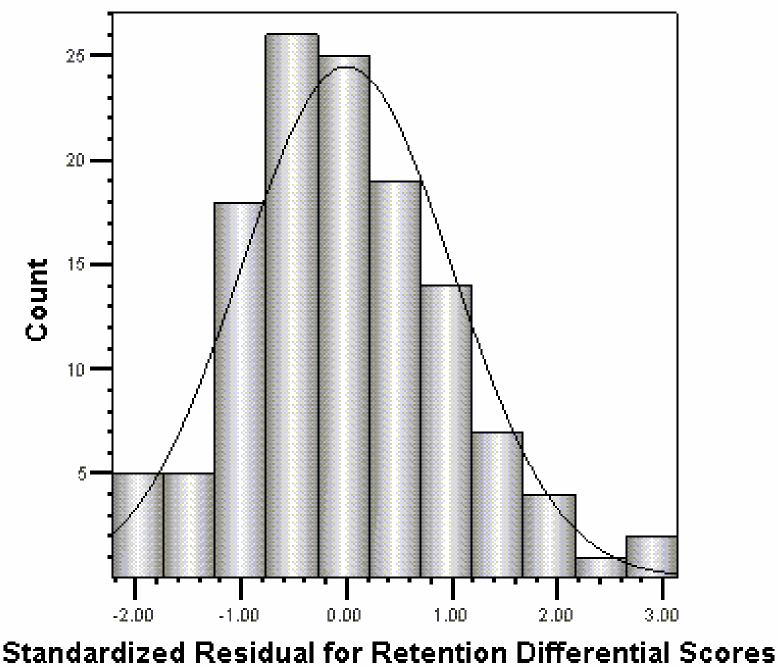


Figure 7.11 Histogram of Standardized Residual for Retention Differential Scores, Fall 2002

To compare the means for differences, an ANOVA was used. The histogram in Figure 7.11 supports the assumption of normality and a Levene's test revealed that there was no difference between the variances ($F(2,123)=0.719$, $p=.489$). With the assumptions met, the ANOVA was conducted and no significant differences were revealed in the mean scores ($F(2,123)=0.243$, $p=.784$). As such, it can be concluded that the differential scores are similar between the three sections investigated.

7.7.2 Spring Retention Data

One hundred and eight students completed the retention test in the spring of 2003, 26 of those were from Section B, 44 from Section E, and 38 were from Section G. These translate into respective return rates of 90, 100, and 84 percents. Three difference of proportions tests revealed two significant difference in return rates (Z_{B-E}=-2.179, $p=.029$; Z_{E-G} =2.726, $p=.007$; Z_{G-B} =-0.640, $p=.522$). The significant differences between the return rate for Section E and the other sections are a result of the 100 percent return rate for Section E. This significance represents only two returned exams (that is to say that if two less exams were returned in Section E, there would be no significant differences at all). The exact implications of this difference are unknown but it is assumed that these two extra tests will not affect the results adversely or otherwise. The analysis was conducted despite these differential return rates.

7.7.2.1 Total Scores on Retention Test

The pretest scores for the spring sections of CEE 2020 are summarized in Table 7.58. While the data did satisfy the normality assumption, as shown in Figure 7.12, the variances were not equal ($F(2,105)=3.115$, $p<.05$). Therefore, as with the fall data, a KW test was conducted to compare the total scores on the retention tests. The KW test revealed that there were no significant differences between the retention test scores for the different sections ($F(2,105)=0.377$, $p=.687$). It can thus be concluded that the sections all performed equally well on the retention tests.

Table 7.58 Descriptive Statistics for Retention Test Scores by Section, Spring 2003

Score on Retention Test

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Section B, Spring 2003	26	24.596	12.313	2.415	19.623	29.570	5.0	48.0
Section E, Spring 2003	44	26.375	7.869	1.186	23.983	28.767	10.0	48.0
Section G, Spring 2003	38	24.829	9.635	1.563	21.662	27.996	2.0	42.5
Total	108	25.403	9.648	.928	23.562	27.243	2.0	48.0

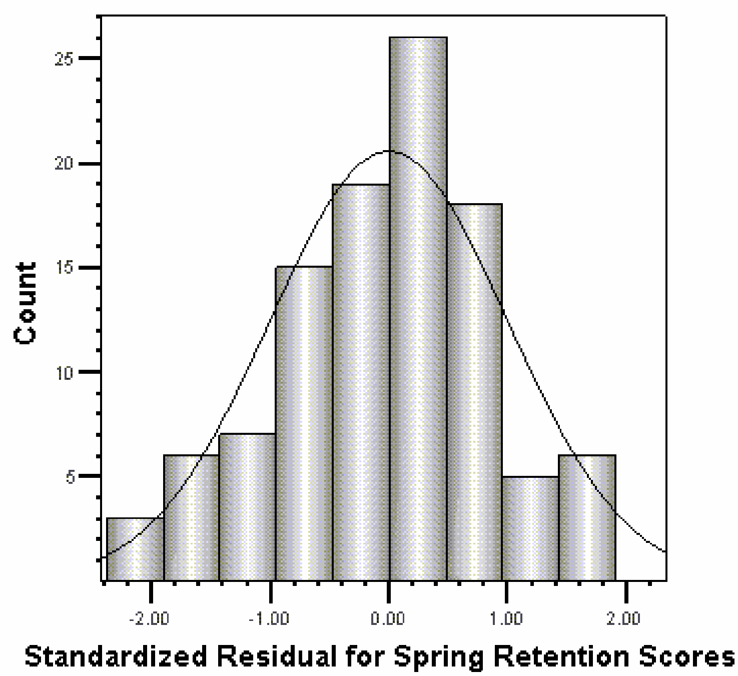


Figure 7.12 Histogram of Standardized Residuals for Scores on the Retention Test, Spring 2003

7.7.2.2 Scores on Individual Question Types

The spring retention data was further analyzed by breaking down the scores according the three different question types. A chi-square test was used for the factual data and KW tests were used for quantitative and qualitative data.

The factual data are presented in Table 7.59. Again, scores of 0.0, 1.0, and 2.0 were all lumped together in order to appropriately use the chi-square test to compare the distribution of scores across the three sections. The test revealed that there were no significant differences between the distributions ($\chi^2(6)=6.888$, p=.331) and as such it can be concluded that each of the sections performed equally well on Problem 1 (i.e. the factual questions) of the retention test.

Table 7.59 Pretest Scores on Factual Questions by Section, Spring 2003

			Score on Problem 1 on Retention Test				Total
			<3.0	3.0	4.0	5.0	
Section ID	Section B, Spring 2003	Count	4	7	13	2	26
		% within Section ID	15.4%	26.9%	50.0%	7.7%	100.0%
	Section E, Spring 2003	Count	5	11	15	13	44
		% within Section ID	11.4%	25.0%	34.1%	29.5%	100.0%
	Section G, Spring 2003	Count	8	9	11	10	38
		% within Section ID	21.1%	23.7%	28.9%	26.3%	100.0%
Total		Count	17	27	39	25	108
		% within Section ID	15.7%	25.0%	36.1%	23.1%	100.0%

The qualitative results are summarized in Table 7.60. The KW test revealed that there were no significant differences in qualitative scores between the three different sections ($\chi^2(2)=1.833$, p=.400). It can be concluded, then, that each of the sections performed equally well on the qualitative portions of the retention test.

Table 7.60 Descriptive Statistics for Qualitative Retention Scores by Section, Spring 2003

Score on Qualitative Retention Test Questions

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Section B, Spring 2003	26	8.5385	5.1399	1.0080	6.4624	10.6145	1.00	19.00
Section E, Spring 2003	44	6.9091	4.4660	.6733	5.5513	8.2669	.00	21.00
Section G, Spring 2003	38	6.5526	3.8252	.6205	5.2953	7.8099	.00	16.00
Total	108	7.1759	4.4571	.4289	6.3257	8.0261	.00	21.00

Table 7.61 summarizes the data for the quantitative questions on the retention test. A KW test was conducted and no significant differences in quantitative scores between the different sections were revealed ($\chi^2(2)=2.156$, p=.340). Once again, it can be concluded that each of the sections performed equally well on the quantitative portion of the retention test.

Table 7.61 Descriptive Statistics for Quantitative Retention Scores by Section, Spring 2003

Score on Quantitative Retention Test Questions

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Section B, Spring 2003	26	12.5962	8.0138	1.5716	9.3593	15.8330	.00	24.00
Section E, Spring 2003	44	15.6477	5.8851	.8872	13.8585	17.4370	.00	24.00
Section G, Spring 2003	38	14.7763	7.3620	1.1943	12.3565	17.1961	.00	24.00
Total	108	14.6065	7.0074	.6743	13.2698	15.9432	.00	24.00

7.7.2.2 Differential Scores

The spring differential retention scores are summarized in Table 7.62. The ranges in the spring semester are slightly higher than those realized in fall semester but the means remained similar and seemed to vary little by section. In order to use the ANOVA to compare these means the assumptions were first assessed. Figure 7.13 reveals that the data were nearly normal and a Levene's test revealed no significant difference between the variances (F(2,96)=2.257, p=.110). With the assumptions satisfied, an ANOVA was conducted, the results of which revealed no significant differences between the differential retention scores (F(2,96)=0.338, p=.714). Once again, it can be concluded that each of sections had similar differential retention scores.

Table 7.62 Descriptive Statistics for Differential Retention Scores, Spring 2003

Retention Differential Score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Section B, Spring 2003	21	4.6667	12.7557	2.7835	-1.1397	10.4730	-16.50	31.50
Section E, Spring 2003	42	2.7500	8.6991	1.3423	3.918E-02	5.4608	-22.50	17.50
Section G, Spring 2003	36	2.3056	11.6553	1.9426	-1.6380	6.2492	-22.00	29.00
Total	99	2.9949	10.6856	1.0739	.8638	5.1261	-22.50	31.50

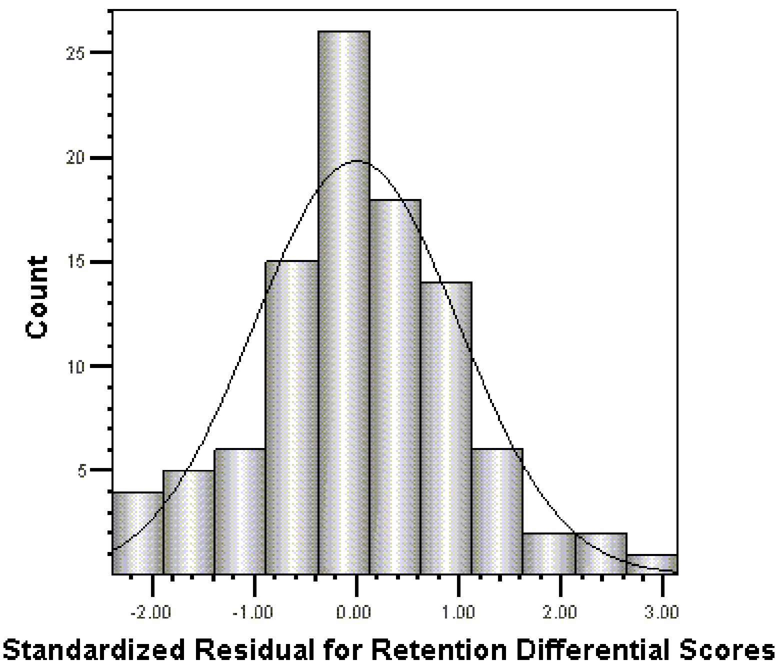


Figure 7.13 Histogram of Standardized Residuals for Differential Retention Scores, Spring 2003

7.7.3 Comparison of Fall and Spring Retention Data

The results of the analyses performed on the fall and spring retention data were identical. They all revealed that there were no significant differences as a result of treatment condition. Because the results were identical in both the initial and replicated studies and because the population distributions were not significantly different in at least one of the studies for both student rank and major of study, there is no reason to assume that there was an effect due to either of these variables. Thus, further investigation into the retention data to determine any effects of rank and major were not completed as was done with the posttest data.

7.7.4 Combined Retention Data

As with the posttest data in Section 7.5.4, the retention test results were lumped across semesters according to treatment condition. Again, this was done purely for illustrative purposes and these analyses should not take precedence over the replicated results as designed and described above. A KW test was performed on the total retention test scores, see Table 7.63, and an ANOVA was performed on the differential retention scores, see Table 7.64, as was done with data from the individual semesters. No significant differences were revealed in either the total scores ($\chi2(2)=3.203$, $p=.202$) or the differential scores ($F(2,222)=0.003$, $p=.997$).

Table 7.63 Descriptive Statistics for Total Score on Retention Test by Treatment Condition

Score on Retention Test

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Content-type Software Group	63	24.397	11.168	1.407	21.584	27.210	3.0	48.0
Tool-type Software Group	81	23.981	9.095	1.011	21.970	25.993	6.0	48.0
Comparison Group	101	21.361	11.421	1.136	19.107	23.616	1.0	45.0
Total	245	23.008	10.687	.683	21.663	24.353	1.0	48.0

Table 7.64 Descriptive Statistics for Differential Retention Scores by Treatment Condition

Retention Differential Score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Content-type Software Group	55	3.0000	10.3199	1.3915	.2101	5.7899	-16.50	31.50
Tool-type Software Group	79	3.0063	8.6438	.9725	1.0702	4.9424	-22.50	21.50
Comparison Group	91	2.9066	10.9158	1.1443	.6333	5.1799	-22.00	32.50
Total	225	2.9644	9.9784	.6652	1.6535	4.2753	-22.50	32.50

7.7.5 Summary of Retention Data

Each of the statistical tests run on the retention data revealed the same result: no significant differences resulting from the treatments. Not only are these results internally consistent (i.e. the results from each individual semester agree) but they are consistent with the posttest results as described in the previous chapter. The implications of these results will be discussed in the next chapter. First, however, the results of the long-term assessments will be presented.

7.8 Long-term Retention Data

The long-term retention test was the final assessment conducted in connection with this research project. The long-term retention test was administered approximately 25 weeks after the posttest and was completed only by CEE students for reasons described previously in this document. The test was administered in CEE 3020, a follow-up class to CEE 2020, where students were given bonus points for completing the test. Students who had participated in the intervention but were not enrolled in CEE 3020 were emailed a copy of the long-term retention test and requested to complete and return the exam on their own accord. As with the retention test, the long-term retention test was a take-home assignment and students were asked to spend no more than 40 minutes on the exam. Students in CEE 3020 were given a week to complete the exam; students who were emailed a copy of the exam were given a few extra days because some were not on campus (e.g. co-ops).

This section will describe the long-term retention results by individual semesters and combined by treatment condition. The results will be discussed in terms of total scores, scores on question types and differential scores, as were the retention results.

7.8.1 Fall Long-term Retention Data

Thirty-five CEE students who participated in the fall implementation of the study completed the long-term retention test. Seven of these were from Section C, which had a total CEE enrollment of 17, yielding a 41 percent return rate. Of the seven who participated from Section C, six of them completed the test in CEE 3020. Two of the students who were enrolled in CEE 3020 did not return the long-term retention test. Nine CEE students from Section C were not enrolled in CEE 3020; these students were sent an email requesting their participation and stressing the importance of their assistance. Only one student contacted via email responded by turning in a completed test. Collecting the results in class was much more effective than collecting them on a one-on-one volunteer basis.

Ten students from Section E, which had a total CEE enrollment of 14, completed the long-term retention test. This translates into a return rate of approximately 71 percent for Section E. Of the ten who completed the test, eight of them participated in CEE 3020 and one student who was enrolled in that follow-up course did not participate. Five students were emailed the exam and two of these returned their completed tests.

Eighteen CEE students from Section H, which had a total CEE enrollment of 29, completed the long-term retention test. The return rate, then was approximately 62 percent. Of the 18 who participated, 16 completed the exam in CEE 3020. Ten students, who were not enrolled in CEE 3020, were emailed the assignment and asked to participate. Of these, two returned their completed tests.

Three difference of proportions tests were completed to compare the return rates from the three sections. These results returned no significant differences (Z_{C-E}=-1.684, p=.090; Z_{E-H} =0.603, p=.546; Z_{H-C} =1.373, p=.170) and it can thus be concluded that an equal percentage of students from each section participated in the long-term retention test.

7.8.1.1 Total Scores

The normality assumption that must be satisfied in order to use parametric statistics becomes less robust with smaller sample sizes and much less so if the samples are of different sizes. Because this was the case with the long-term retention data, nonparametric statistics were used. The results from the fall sections long-term assessment are presented in Table 7.65. A KW test was used to compare the scores; the test revealed no significant differences ($\chi 2(2)$=0.529, p=.768).

Table 7.65 Descriptive Statistics for Long-term Retention Scores by Section, Fall 2002

Score on Long-term Retention Test

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Section C, Fall 2002	7	22.571	12.153	4.594	11.332	33.811	7.0	40.0
Section E, Fall 2002	10	23.800	10.441	3.302	16.331	31.269	8.0	38.5
Section H, Fall 2002	18	20.972	10.538	2.484	15.732	26.213	6.0	40.0
Total	35	22.100	10.584	1.789	18.464	25.736	6.0	40.0

7.8.1.2 Scores on Individual Question Types

As with the analyses of the posttest and the retention test, long-term retention scores were broken down into scores on individual question types: factual, quantitative and qualitative. Unfortunately, the sample size was insufficient to allow for a chi-square comparison of the factual questions.

The quantitative questions were compared using the KW test, as was done previously. The data are summarized in Table 7.66, and though Section E has a higher mean than the others, the KW test revealed no significant differences ($\chi 2(2)=0.489$, $p=.783$). The qualitative data are presented in Table 7.67. A KW test revealed that the qualitative scores were not significantly different ($\chi 2(2)=1.154$, $p=.562$). It can thus be concluded that each of the sections performed equally well on both qualitative and quantitative questions.

Table 7.66 Descriptive Statistics for Quantitative Long-term Retention Scores, Fall 2002

Score on Quantitative Long-term Retention Test Questions

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Section C, Fall 2002	7	11.4286	8.8149	3.3317	3.2762	19.5810	.00	23.00
Section E, Fall 2002	10	13.6000	7.2641	2.2971	8.4036	18.7964	.00	22.50
Section H, Fall 2002	18	11.6389	7.5592	1.7817	7.8798	15.3980	.00	23.00
Total	35	12.1571	7.5574	1.2774	9.5611	14.7532	.00	23.00

Table 7.67 Descriptive Statistics for Qualitative Long-term Retention Scores, Fall 2002

Score on Qualitative Long-term Retention Test Questions

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Section C, Fall 2002	7	6.5714	5.1594	1.9501	1.7998	11.3430	2.00	16.00
Section E, Fall 2002	10	6.9000	4.2282	1.3371	3.8753	9.9247	1.00	15.00
Section H, Fall 2002	18	5.2778	3.6591	.8625	3.4582	7.0974	1.00	12.00
Total	35	6.0000	4.0873	.6909	4.5960	7.4040	1.00	16.00

7.8.1.3 Long-term Differential Scores

Nonparametric statistics were again used to compare the differential scores because of low and unequal samples sizes. The fall long-term differential scores are summarized in Table 7.68. Note that despite scores ranging from moderately negative to moderately positive (recall that the potential range is from –55 to +55), the means revealed small to no increase, but none of the means were negative. This implies that, on average, students do indeed remember what is taught them, even after 25 weeks—well into the subsequent semester when students are expected to apply what they have learned to new and different topics in other courses. A KW test revealed no significant differences in the differential scores ($\chi2(2)=0.367$, $p=.832$). As such, it can be concluded that the sections each had similar differential scores, that each section recalled the same amount of information, on average.

Table 7.68 Descriptive Statistics for Long-term Differential Scores, Fall 2002

Long-term Differential Scores

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Section C, Fall 2002	6	.0000	10.8167	4.4159	-11.3514	11.3514	-18.50	8.50
Section E, Fall 2002	10	6.6000	13.3079	4.2083	-2.9199	16.1199	-9.00	27.00
Section H, Fall 2002	18	1.2500	11.8163	2.7851	-4.6261	7.1261	-21.00	24.50
Total	34	2.6029	12.0427	2.0653	-1.5990	6.8048	-21.00	27.00

7.8.2 Spring Long-term Retention Data

Twenty-five CEE students who participated in the fall implementation of the study completed the long-term retention test. Seven of these were from Section B, which had a total CEE enrollment of 16, yielding a 44 percent return rate. Of the seven who participated from Section B, five of them completed the test in CEE 3020. Nine CEE students from Section B were not enrolled in CEE 3020; these students were sent an email requesting their participation and stressing the importance of their assistance. Two of the student contacted via email responded by turning in their completed tests. Once again, collecting the results in class was much more effective than collecting them on a one-on-one volunteer basis via email.

Thirteen CEE students from Section E, which had a total CEE enrollment of 21, completed the long-term retention test. This translates into a return rate of approximately 62 percent for Section E. Of the 13 who completed the test, nine of them participated in CEE 3020 and two students who were enrolled in that follow-up course did not participate. Ten students were emailed the exam and four of these returned their completed tests.

Five CEE students from Section G, which had a total CEE enrollment of nine, completed the long-term retention test. The return rate was approximately 56 percent. Of the five who participated, three completed the exam in CEE 3020. Six students, who were not enrolled in CEE 3020, were emailed the assignment and asked to participate. Of these, two returned their completed tests.

Three difference of proportions tests were completed to compare the return rates from the three sections. These results returned no significant differences (Z_{B-E}=-1.098, p=.272; Z_{E-G} =0.325, p=.775; Z_{G-B} =0.567, p=.571) and it can thus be concluded that an equal percentage of students from each section participated in the long-term retention test.

7.8.2.1 Total Scores

The sample sizes during the spring semester were once again quite small and unequal. As such, the robustness of the normality assumption could not be relied upon and as such nonparametric statistics were used. Table 7.69 presents the total long-term retention scores for the spring sections. A KW test was used to compare these scores and no significant differences were found ($\chi^2(2)$=2.430, p=.297).

Table 7.69 Descriptive Statistics for Total Long-term Retention Scores, Spring 2003

Score on Long-term Retention Test

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Section B, Spring 2003	7	32.786	6.383	2.412	26.883	38.689	23.5	41.0
Section E, Spring 2003	13	25.423	11.565	3.207	18.435	32.412	4.0	40.0
Section G, Spring 2003	5	25.100	13.297	5.946	8.590	41.610	12.0	46.0
Total	25	27.420	10.872	2.174	22.932	31.908	4.0	46.0

7.8.2.2 Scores on Individual Question Types

Again, there were too few students in the sample groups to successfully and appropriately complete a chi-square comparison of the factual scores. The quantitative scores, presented in table 7.70, were compared across sections using the KW test. The test revealed that there were no significant differences ($\chi^2(2)=1.804$, $p=.406$) between any of the groups in terms of quantitative scores on the long-term retention test. Similarly, the qualitative scores, presented in Table 7.71, were also compared and no significant differences were revealed ($\chi^2(2)=1.937$, $p=.380$). It can thus be concluded that each of the sections performed equally well on the qualitative and the quantitative questions and that any differences in scores are due strictly to chance.

Table 7.70 Descriptive Statistics for Scores on the Long-term Retention Quantitative Questions, Spring 2003

Score on Quantitative Long-term Retention Test Questions

					95% Confidence Interval for Mean			
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Section B, Spring 2003	7	18.3571	4.8280	1.8248	13.8920	22.8223	11.50	24.00
Section E, Spring 2003	13	14.3462	7.4649	2.0704	9.8352	18.8571	.00	23.00
Section G, Spring 2003	5	14.3000	7.2250	3.2311	5.3290	23.2710	4.00	22.00
Total	25	15.4600	6.7668	1.3534	12.6668	18.2532	.00	24.00

Table 7.71 Descriptive Statistics for Scores on the Long-term Retention Qualitative Questions, Spring 2003

Score on Qualitative Long-term Retention Test Questions

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Section B, Spring 2003	7	9.7143	3.4983	1.3222	6.4789	12.9497	5.00	16.00
Section E, Spring 2003	13	6.8462	5.3361	1.4800	3.6216	10.0707	.00	15.00
Section G, Spring 2003	5	6.8000	7.2938	3.2619	-2.2565	15.8565	1.00	19.00
Total	25	7.6400	5.2827	1.0565	5.4594	9.8206	.00	19.00

7.8.2.3 Long-term Differential Scores

The spring semester long-term differential scores, or difference between the scores on the long-term retention test and scores on the posttest, are presented in Table 7.72. Again, the means were all positive suggesting more was remembered than was forgot, on average. A KW test was used to compare the scores and no significant difference was found ($\chi^2(2)=0.545$, $p=.761$).

Table 7.72 Descriptive Statistics for Long-term Differential Scores, Spring 2003

Long-term Differential Scores

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Section B, Spring 2003	6	5.5000	10.0150	4.0886	-5.0101	16.0101	-9.50	17.00
Section E, Spring 2003	13	2.2692	7.7207	2.1413	-2.3963	6.9348	-12.00	16.00
Section G, Spring 2003	5	2.4000	8.5980	3.8451	-8.2758	13.0758	-9.00	15.00
Total	24	3.1042	8.2317	1.6803	-.3718	6.5801	-12.00	17.00

7.8.3 Comparison of Fall and Spring Data

The long-term retention results were consistent across semesters. All statistical tests revealed no significant differences as a result of the intervention or treatment condition. It can be concluded from these results that the sections performed equally well on the retention test as a whole and on the individual question types on the retention test. Also, each of the sections realized the same minor improvement in truss knowledge and abilities over the twenty-five week period.

7.8.4 Combination of Fall and Spring Data

As was done previously, groups from individual semesters were combined according to the treatment condition they participated in during the intervention. The combined results, as mentioned previously, are presented here for illustrative purposes. This research was designed to be based upon replicated results from individual semesters and not upon combined data from two different semesters and thus the replicated results take precedence over the combined results. Combined total scores, as summarized in Table 7.73, were compared using a KW test; the test revealed no significant differences resulting from treatment condition ($\chi^2(2)=2.772$, $p=.250$). The differential scores, summarized in Table 7.74, were compared as before with an ANOVA, which also revealed no significant differences ($F(2,55)=0.355$, $p=.703$).

Table 7.73 Descriptive Statistics for Scores on the Long-term Retention Test, All Sections

Score on Long-term Retention Test

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Content-type Software Group	14	27.679	10.727	2.867	21.485	33.872	7.0	41.0
Tool-type Software Group	23	24.717	10.873	2.267	20.016	29.419	4.0	40.0
Comparison Group	23	21.870	10.999	2.293	17.113	26.626	6.0	46.0
Total	60	24.317	10.937	1.412	21.491	27.142	4.0	46.0

Table 7.74 Descriptive Statistics for Differential Long-term Retention Scores, All Sections

Long-term Differential Scores

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Content-type Software Group	12	2.7500	10.3452	2.9864	-3.8230	9.3230	-18.50	17.00
Tool-type Software Group	23	4.1522	10.4777	2.1848	-.3787	8.6831	-12.00	27.00
Comparison Group	23	1.5000	11.0258	2.2990	-3.2679	6.2679	-21.00	24.50
Total	58	2.8103	10.5531	1.3857	3.556E-02	5.5851	-21.00	27.00

7.8.5 Summary of Long-term Retention Data

The results from the long-term retention analyses were internally and externally consistent. They were internally consistent because each of the tests performed on the long-term retention data yielded the same results: no significant difference as a result of treatment condition. External consistency comes because these results are also identical to the results from posttest, exam questions, and retention data, all of these revealed no significant differences. The implications of these results will be presented in the following chapter.

7.9 Scores Across Time

An inspection of the differential scores prompted one additional analysis. It was assumed, based on the literature, that the average differential scores would be negative. In other words, it was assumed that students would forget truss analysis over time and thus perform more poorly on the exams as time went on. The differential scores, however, indicate differently. The average differential scores were all positive, with values in the range of three to four points. These data suggest that students not only retained what was being taught, but may have gained some degree of knowledge in the period of time between the posttest and the other tests. To determine if an increase actually did take place, an analysis of scores across time was completed.

An ANOVA was used to simultaneously compare the total scores on the posttest, retention test, and long-term retention test to determine if the scores changed over time. The ANOVA assumptions were first tested and while the data were not normal, the distribution was quite symmetric (see Figure 7.14) and the normality assumption was thus satisfied. Furthermore, it was determined by way of a Levene's test that there was no significant difference between the variances ($F(2,545)=1.856$, $p=.157$). The omnibus test revealed that there was a significant difference between the scores over time ($F(2,545)=4.992$, $p<.05$). Post-hoc comparisons via the Tukey method revealed that there was a significant difference between the mean scores for the retention test and the posttest (2.405, $p<.05$) and there was a significant difference between the mean scores for the long-term retention test and the posttest (3.714, $p<.05$). There was, however, no significant difference between the long-term retention and retention test scores (1.309, $p<.05$). These results indicate that students did gain some knowledge between the posttest and the retention test but that the degree of their knowledge remained somewhat constant after that point. The possible causes for and implications of these results will be discussed in the following chapter.

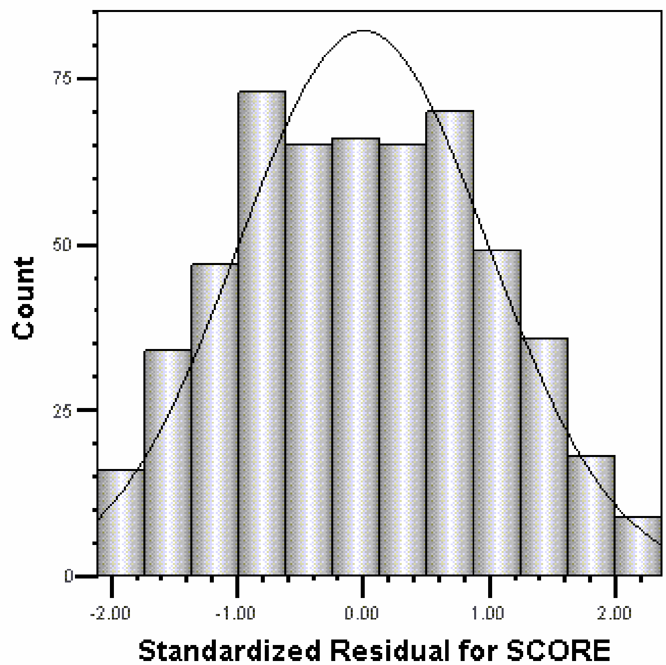


Figure 7.14 Histogram of Standardized Residuals for Total Scores on Posttest, Retention Test, and Long-term Retention Test

7.10 Survey Questions

In addition to the demographic questions asked during the gathering of the retention data, the experimental groups were also asked six questions about their experience with the software. A fully developed and tested survey was not used in this study because the purpose of this project was to determine the effects of software use learning and retention, not to investigate the *perceived* benefits of software use. Students' opinions, attitudes, and perceptions have been studied elsewhere (e.g. Riley and Pace, 1996). These questions were asked simply to get the students' opinions of the specific software titles that they used. Following are the questions that students were asked to which they responded with a rating on a scale of one to five, where 1 = No, 2 = Not Really, 3 = Uncertain, 4 = Somewhat, and 5 = Yes.

- Did you enjoy using the software?
- Did you use the software for topics other than trusses?
- Would you use the software again?
- Are you still using the software?
- Would you recommend using the software in future classes?
- Do you think the software helped you understand trusses better?

The results of these informal surveys are presented in Table 7.75, 7.76, 7.77, and 7.78 for sections C from fall, E from fall, B from spring, and E from spring respectively. The tables reveal that most students somewhat enjoyed using the software, though Fall Section C was uncertain. Almost all students had not used the software for other topics nor were they still using it. Both fall sections were uncertain whether they would ever use the software again; this is understandable considering students may not know for sure whether the software would be applicable for the content of future courses. The spring sections replied, on average, that they might use the software again. All sections agreed that they would somewhat recommend using the software in future courses. This is a good sign and suggests that they did perceive some benefit from use of the software. There was less agreement, however, as to whether the software actually helped the students understand trusses: the fall sections were uncertain while the spring sections answered somewhat. In this case, the quantitative results support the students' perceptions; the use of the software did not significantly increase learning or retention.

7.11 Summary of Data Analysis

In nearly all of the statistical tests conducted, there were no significant differences between any score or measure as a result of the treatment condition. In the few cases where there were significant differences (on some of the individual question types on the posttest), these differences were not replicated and were thus not considered valid. The conclusions and implications of these results will be discussed in the following chapter.

Table 7.75 Frequencies of Student Responses to Survey Questions, Fall Section C, Content-type Software

Rank	Enjoyed using the software	Used software for other topics	Would use software again	Still using software	Would recommend using software in future classes	Software use helped understanding of trusses
(1) No	6 (16.2%)	31 (83.8%)	13 (35.1%)	30 (81.1%)	5 (13.5%)	10 (27.8%)
(2) Not Really	4 (27.0%)	3 (91.9%)	2 (40.5%)	4 (91.9%)	3 (21.6%)	2 (33.3%)
(3) Uncertain	11 (56.8%)	1 (94.6%)	13 (75.7%)	3 (100.0%)	8 (43.2%)	10 (61.1%)
(4) Somewhat	13 (91.9%)	0 (94.6%)	7 (94.6%)	0 (100.0%)	15 (83.8%)	8 (83.3%)
(5) Yes	3 (100.0%)	2 (100.0%)	2 (100.0%)	0 (100.0%)	6 (100.0%)	6 (100.0%)
Median	3 (Uncertain)	1 (No)	3 (Uncertain)	1 (No)	4 (Somewhat)	3 (Uncertain)

Table 7.76 Frequencies of Student Responses to Survey Questions, Fall Section E, Tool-type Software

Rank	Enjoyed using the software	Used software for other topics	Would use software again	Still using software	Would recommend using software in future classes	Software use helped understanding of trusses
(1) No	0 (0.0%)	35 (94.6%)	6 (16.2%)	29 (78.4%)	1 (2.7%)	3 (8.1%)
(2) Not Really	1 (2.8%)	2 (100.0%)	4 (27.0%)	6 (94.6%)	1 (5.4%)	4 (18.9%)
(3) Uncertain	10 (30.6%)	0 (100.0%)	10 (54.1%)	1 (97.3%)	5 (18.9%)	14 (56.8%)
(4) Somewhat	13 (66.7%)	0 (100.0%)	9 (78.4%)	1 (100.0%)	19 (70.3%)	9 (81.1%)
(5) Yes	12 (100.0%)	0 (100.0%)	8 (100.0%)	0 (100.0%)	11 (100.0%)	7 (100.0%)
Median	4 (Somewhat)	1 (No)	3 (Uncertain)	1 (No)	4 (Somewhat)	3 (Uncertain)

Table 7.77 Frequencies of Student Responses to Survey Questions, Spring Section B, Content-type Software

Rank	Enjoyed using the software	Used software for other topics	Would use software again	Still using software	Would recommend using software in future classes	Software use helped understanding of trusses
(1) No	2 (8.3%)	18 (72.0%)	7 (28.0%)	20 (80.0%)	1 (4.2%)	2 (8.3%)
(2) Not Really	0 (8.3%)	1 (76.0%)	1 (32.0%)	2 (88.0%)	0 (4.2%)	0 (8.3%)
(3) Uncertain	4 (25.0%)	4 (92.0%)	4 (48.0%)	1 (92.0%)	6 (29.2%)	5 (29.2%)
(4) Somewhat	9 (62.5%)	1 (96.0%)	5 (68.0%)	0 (92.0%)	8 (62.5%)	8 (62.5%)
(5) Yes	9 (100.0%)	1 (100.0%)	8 (100.0%)	2 (100.0%)	9 (100.0%)	9 (100.0%)
Median	4 (Somewhat)	1 (No)	4 (Somewhat)	1 (No)	4 (Somewhat)	4 (Somewhat)

Table 7.78 Frequencies of Student Responses to Survey Questions, Spring Section E, Tool-type Software

Rank	Enjoyed using the software	Used software for other topics	Would use software again	Still using software	Would recommend using software in future classes	Software use helped understanding of trusses
(1) No	1 (2.3%)	40 (90.9%)	5 (11.4%)	38 (86.4%)	1 (2.3%)	5 (11.4%)
(2) Not Really	1 (4.5%)	2 (95.5%)	5 (22.7%)	4 (95.5%)	1 (4.5%)	3 (18.2%)
(3) Uncertain	10 (27.3%)	1 (97.7%)	7 (38.6%)	0 (95.5%)	8 (22.7%)	6 (31.8%)
(4) Somewhat	16 (63.6%)	1 (100.0%)	10 (61.4%)	2 (100.0%)	18 (63.6%)	16 (68.2%)
(5) Yes	16 (100.0%)	0 (100.0%)	17 (100.0%)	0 (100.0%)	16 (100.0%)	14 (100.0%)
Median	4 (Somewhat)	1 (No)	4 (Somewhat)	1 (No)	4 (Somewhat)	4 (Somewhat)

CHAPTER 8

FINDINGS

This chapter will discuss the implications of the numerical results presented in the previous chapter. Also, possible reasons for the results will be presented and discussed. The chapter will begin by discussing the necessity of the pretest, the pretest results, and the implications of these results on the further findings. The chapter will continue to be divided by assessment instruments, each of which was driven by two hypotheses. These hypotheses will be restated and conclusions will be drawn about each of them based on the results. Finally, the findings will be compared to other research and conclusions about the findings will be presented.

8.1 Pretest and Demographic Findings

Gathering pretest and demographic data was an essential part of the research design. The reason for doing this was to remove the possibility that an undesirable effect would result from a non-experimental variable such as gender, rank, or prior knowledge. This was especially of importance in this research project because it followed a quasi-experimental design, which is to say that the participants were not randomly selected or assigned to sample groups. As such, the non-experimental effects could not be assumed to be random either. Entire sections of CEE 2020 were used as sample groups and it was possible that the non-experimental variables, such as friends or classmates of similar rank and major, influenced a student or students to take a particular section and thus be included in one of the sample groups. Thus, demographic and prior knowledge data needed to be accounted for prior to making any meaningful conclusions about learning and retention.

Some of the demographic information was gathered from the class rolls while other information was based on student responses to a demographic questionnaire that was handed out with the retention test. Prior knowledge was gauged through the use of a pretest that assessed knowledge that students must have mastered in order to analyze trusses correctly.

The analysis techniques described in the previous chapters were used to compare these non-experimental variables across the sections or treatment conditions to determine if the sections differed significantly in any of the areas. These comparisons were conducted for both implementations of the research as designed. Parametric and nonparametric techniques were used in the analysis, all at the five percent significance level (a=.05).

The most notable difference in non-experimental variables between treatment conditions occurred in the fall semester of 2002 where the control group had significantly more seniors than was expected. Potentially, if seniors performed differently on the assessment measures than other students, the results of those assessments could be biased as a result of this inequality of student rank. As described in Hays (1996) the best way to combat a potential selection bias is by replicating the experiment. In the replicated study, there was no difference in the distributions of students by rank between the three participating sections. Thus, any results that were realized in the fall but not in the spring could have been caused by a selection bias. Fortunately, nearly all the results were identical in both the original and replicated study and it can thus be assumed that student rank had no effect on any of the assessments.

A similar situation occurred with the major variable. In the case of majors, however, the significant differences occurred in the spring. Specifically, the control group had significantly more IE students than expected. Again, replicated results revealed that any sampling bias that may have led to an inequality of majors did not significantly affect the results of any of the assessments.

There was no other significant differences between any of the demographic data. Nor were there any differences in pretest scores. The pretest findings can then be summarized as follows:

- In both implementations of the study, there were no significant differences in prior-knowledge, as measured by the pretest, between the participating sections.
- In both implementations of the study, there were no significant differences in most of the demographic variables including GPA, gender, ethnicity, and number of credits taken during the intervention.
- In the fall implementation, the three sections did differ significantly in their distributions of student rank. In the spring, there was a significant difference in terms of major. If the results of any of the assessments varied from semester to semester (or in subsequent implementations) then major and rank would have to be included as factors in the analysis to determine the effects of these deviations from equality.

These results were ideal because they allowed the research to continue without any concern as to how to account for nuisance variables. It was assumed after this analysis, then, that the only difference between the three groups each semester was the treatment condition that they experienced. Because there were slight differences, due to rank in one semester and major in the other, only results that were significant in both semesters would be considered valid. If results were isolated to a single semester, further investigation could have been undertaken to include the non-experimental factors. As will be seen shortly, however, it never became necessary to use additional factors in the analysis because, with a few minor exceptions, the results were the same in both implementations of the study.

8.2 Effects of Technology on Learning

Learning was defined earlier in this document as the amount of knowledge a student recalled immediately after the interventions. It was further defined as how well a student performs on a posttest completed in the lecture period immediately following the special session of the intervention. It was hypothesized that the use of technology could enhance learning, or more specifically increase students' posttest scores. More formally, it was hypothesized that:

1. Students who used technology, whether tool-type or content-type, would perform better on the posttest than students who did not use technology.
2. Students who used the tool-type software would perform equally as well on the posttest as students who used the content-type software.

The first hypothesis was tested through the use of a planned comparison that tested whether the average of the two experimental group means was equal to the control group mean. The test revealed that there were was a significant difference between the experimental groups and the control group in the fall study. The second hypothesis was also tested with a planned contrast that revealed that the content group performed better than the tool-type group in the fall study. This was in contrast to the spring study, which revealed no significant differences in either of the planned contrasts.

It was theorized that the differences in the fall study may have been a result of the differences in rank during the fall study, recall that one section had significantly more seniors than the other. To test this theory, a more precise, two-way ANOVA was conducted with treatment condition and rank as factors. This factorial analysis removed some of the variability from the treatment condition and factor and revealed that there were no significant differences between the posttest scores of any of the fall groups. These results were consistent with the spring results. As such, the first hypothesis was rejected and the second was retained.

1. Hypothesis 1: Rejected
2. Hypothesis 2: Not rejected.

Furthermore, because the experimental groups were equal and because the average of the experimental groups was equal to the control group, it can be concluded that all the groups were equal. Thus, it can be concluded that all groups performed equally well on the posttest. The ultimate conclusion that can be drawn from these results is that technology had no effect, either positive or negative, on learning as measured by the posttest.

When selecting the software to use in the intervention, a group of CEE 2020 instructors agreed that the students should be able to manipulate trusses within the software environment and immediately see the results of this manipulation (see Chapter 4 for more information). Based upon this input, the software was chosen and the special session was designed to support these kinds of activities; where students manipulated trusses without necessarily solving them quantitatively. It was assumed, then, that students who used the software would be better prepared at solving these qualitative type questions on the posttest. To test this assumption, the scores on the posttest were broken down by question type and compared across treatment conditions. The three question types tested were as follows:

- Factual: Recall of truss facts and assumptions
- Quantitative: Quantitative analysis of trusses and truss members
- Qualitative: Manipulation of trusses and members to determine effects without solving the truss quantitatively, as described above

When the scores on these questions types were compared between the three groups, only one significant relationship existed: in the fall implementation, the content-type group performed better than the comparison group on the quantitative questions. This was unexpected because, as described above, it was assumed that the students who used the software would be better prepared to solve qualitative questions but there was no reason to assume that using the software would better prepare them for the quantitative problems. Ultimately, however, the reason for this phenomenon was not investigated because the results were not replicated in the subsequent semester. As such, this finding was not considered valid. With regards to the individual question types, it was concluded that there was no significant differences in the scores resulting from the treatment conditions. In other words, using the software did not have an effect on how students performed on individual posttest question types.

One other measure of learning was the exam questions that students completed during each implementation of the study. The exam questions were on midterm exams that students completed approximately one week after the intervention. Students within each semester completed the same exam questions though different questions were used in the spring and fall semesters. Two of the formally stated hypotheses referred to the exam questions; they are as follows:

3. Students who used technology, whether tool-type or content-type, would perform better on the exam questions than students who did not use technology.
4. Student who used the tool-type software would perform equally as well on the exam questions as students who used the content-type software.

To test these, an omnibus hypothesis was tested which compared the exam scores for each of the treatment conditions from each semester. The omnibus test revealed no significant differences in both the fall and the spring. As such, it was concluded each group of students performed equally well on the exam questions, effectively rejecting the third hypothesis and retaining the forth.

Hypothesis 3: Rejected

Hypothesis 4: Not rejected

This further supports the conclusion that software use did not have any effect on learning.

8.4 Effects of Technology on Retention

Within the scope of this research, retention was defined as the amount of information a student recalled ten weeks after the intervention. Retention was further defined as how well students performed on the retention test that was administered approximately ten weeks after the intervention. It was hypothesized that using technology would improve retention. More specifically, the following two hypotheses related to retention:

5. Students who used technology, whether tool-type or content-type, would perform better on the retention test than students who did not use technology.
6. Student who used the tool-type software would perform equally as well on the retention test as students who used the content-type software.

Although it would have been ideal to test the hypotheses with planned comparisons, as the posttest hypotheses were, the retention data were not normal and so the t test used in planned comparisons would not be appropriate. Instead, an omnibus hypothesis was tested, one which compared the scores between all three sections at once. The omnibus hypothesis revealed no differences in scores in either the spring or the fall semesters. Because there were no differences between the experimental group scores and the control group scores, the first hypothesis was rejected. Because there was no difference between the content-type group scores and the tool-type group scores, the second hypothesis could not be rejected.

Hypothesis 5: Rejected

Hypothesis 6: Not rejected.

These results suggest that the students performed equally well on the retention test regardless of the group to which they belonged. In other words, technology use had no effect on retention, either positive or negative.

As with the posttest scores, the retention data were broken down by question type to determine if technology helped students retain more information about specific questions. When comparisons were made, there were no significant differences in scores on any of the question types resulting from the treatment in either fall or spring. It was thus concluded that the use of technology did not have an effect on how students performed on the individual question types on the retention test.

As an additional measure of retention, an analysis of differential scores was conducted. Differential scores were calculated by subtracting each student's posttest score from their respective retention test scores. Differential scores revealed how much information or ability students actually retained over time. The differential scores were compared across the sections in each of the semesters in which the study took place. Each of these comparisons had the same result: no significant difference. Software use did not have an effect on differential scores. It can be concluded, based on these findings, that software use did not help students retain information, nor did it hinder their retention. Student knowledge retention was not affected by technology use.

8.5 Effects of Technology on Long-term Retention

Within the context of this project, long-term retention was defined as the amount of information students recalled 25 weeks after the intervention. More specifically, it was defined as how well students performed on a long-term retention test administered approximately 25 weeks after the intervention. This test was identical to the retention test, which was also identical to the posttest. It was hypothesized that technology use would improve long-term retention. More formally, it was hypothesized that:

7. Students who used technology, whether tool-type or content-type, would perform better on the long-term retention test than students who did not use technology.
8. Student who used the tool-type software would perform equally as well on the long-term retention test as students who used the content-type software

Again, these hypotheses were actually tested simultaneously using an omnibus test that revealed no significant difference in total scores as a result of treatment condition in either the fall or the spring. It can be concluded, then, that each of the groups performed equally well on the long-term retention score, which resulted in the following findings:

Hypothesis 7: Rejected

Hypothesis 8: Not rejected

These findings support the conclusion that technology use had no effect on long-term knowledge retention.

Again, scores were broken down by question type to determine if technology use had an effect on the types of questions students were able to answer over time. Comparisons across semesters revealed no significant differences in either the fall or spring study. It was concluded, therefore, that software use did not have an effect on how students performed on individual questions types.

Differential scores were again analyzed. In regards to long-term retention, differential scores were the difference between a students score on the long-term retention test and the posttest. A comparison of the differential scores across treatment conditions within each semester revealed no significant differences in either fall or spring. Based on these findings, it was concluded that technology use did not help or hinder students' long-term retention. Students who used the software realized the same degree of long-term retention as students who completed problems by hand.

8.6 Effects Over Time

While no formal hypotheses addressed the degree of students' retention over time, the data suggested that an investigation into this matter would be illustrative. It was assumed that the students' knowledge and abilities would wane over time or in other words that they would forget. It was further assumed that they would perform more poorly on the exams as time progressed as result of this forgetting. The data, however, implied that students retained much of what was taught as evidenced by slight positive average differential scores. These positive differentials suggested that students not only retained what was taught but may have picked up additional knowledge in the interim period. Total scores from the three different exams were compared over time to see how they related and there was a significant difference between the posttest and the two retention tests; there was however, no difference between the retention test and the long-term retention test. These results support the finding that students increased in knowledge and ability at some point in time between the posttest and the retention test, after which time their knowledge level seems to have remained constant.

Two conclusions arise from these results: first, students do retain knowledge of trusses and truss analysis as evidenced by the fact that their scores before and after the fifteen week period between retention tests were similar. Second, students in these courses gained additional knowledge in the ten-week period between the posttest and retention test. Possible sources of this additional knowledge will be discussed in the following section.

8.7 Discussion of Findings

Many of the findings revealed in this study were unexpected, or were contrary to the theorized link between the instructional technology and retention. This section attempts to explain the findings and compare them to work by other researchers. Prior to this explanation, however, the two most important findings from this study are summarized.

8.7.1 Discussion of Main Findings

The main objective of this research was to determine if software use would increase knowledge retention. The results suggested the following two main findings:

- Students who used software did not learn or retain more or less information than students who did not use the software, all groups performed equally well on all assessments.
- Students in each of the groups did retain information across time. Mean scores on each subsequent assessment were at least as good as those on the previous assessments.

In other words, students did retain a significant amount of information, but their retention rates did not differ as a result of the treatment they experienced.

The reasons for these unexpected results cannot be determined from the data but possible reasons can be inferred from work done by others. It was assumed that all groups performed equally well because the special session was not a passive experience for the control group students. In the special session, the control groups did not passively attend a lecture, as in other studies (e.g. Riley and Pace, 1997). In the special session, control-group students broke into pairs and solved truss analysis questions by hand. This was obviously an active process, though the activity differed in nature from the activity that the experimental groups were involved in. Additionally, this was a form of alternative instruction because the normal instructors for both of the control groups had never used class time in this manner.

It was assumed, then, that software use itself did not affect retention, but that it was either the active or alternative nature of the instruction that kept retention rates high in both the experimental and control groups. Which of these two variables, activity or alternative instruction, is actually the cause for the increased retention is unknown. The effects of these two variables cannot be separated in the data from this experiment but another experiment could possibly be designed to control for activity and/or alternative instruction.

When compared to other studies on the effects of IT, these findings are consistent with some findings but in contrast with others. Sulbaran, for example, (2002) found that students who received virtual reality instruction did not learn or retain more information in the short term (one week) than students who learned via static web materials or traditional lectures. Alternatively, Issa et al. (1999) found that students who received multimedia instruction via a CD-ROM learned and retained more information in the short term (three weeks) than students who were instructed in a traditional lecture style. These results are similar to those found by Kulik and Kulik (1991) who analyzed a number of IT studies, not for retention but for learning, and found that in almost all cases IT was at least as good as traditional instruction. Unfortunately, Kulik and Kulik had no explanations as to why some studies led to increased learning and others did not.

Laboratory studies have revealed similar results. Mayer (2001) conducted a number of studies on computer-based multimedia and learning in controlled laboratory studies and found that in some cases subjects who used multimedia had higher retention rates—assessed after retention intervals of a few minutes—than students who did not. In other studies, subjects in all studies performed the same, but in no studies did the multimedia subjects perform worse than the control subjects. This again suggests that multimedia or IT is at least as good as traditional instruction. Mayer (2001) was unable to explain why this is the case and called for more studies so that a link could be established between retention and multimedia.

As stated at the beginning of this document, however, this research was intended to be practical in nature and not theoretical. The purpose of this research was not to answer the *why* question, the purpose was to determine *what* the long-term effects of technology use were. As the results show, technology use is at least as good as traditional instruction and students were no worse off for using it. In fact, students did retain much more information than expected. Unexpectedly, however, the control group did as well. These findings have some practical implications for instructors who are considering IT use in their courses; these implications will be discussed in the Section 8.8.

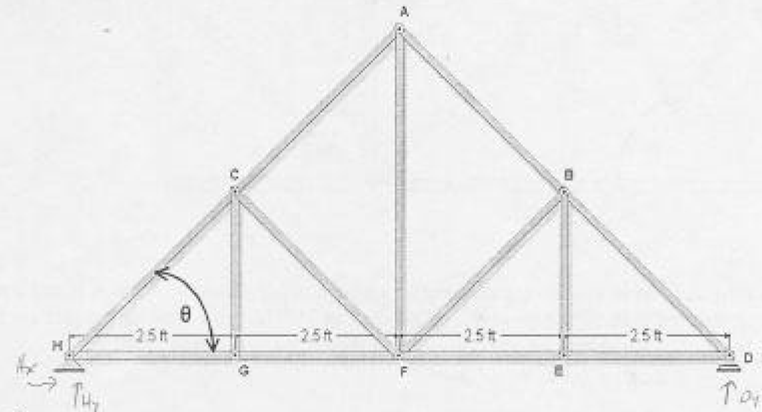
8.7.2 Discussion on Efficiency

Though the findings reveal that software use had no effects on learning and retention, there was one area where software use did have an impact: problem-solving efficiency. Efficiency was not measured quantitatively, but software use had an obvious effect. Students who used software during the special session were able to complete nearly all of the assigned exercises whereas students who worked out the problems by hand completed only a few. This is illustrated by the examples given in Figures 8.1 and 8.2, which show completed special session exercises for a student in the control group and a student in the content-type group respectively (tool-type students had similar results to the content-type students). While some students from each group may have completed more or less than each of the examples given here, exercises shown in Figures 8.1 and 8.2 are typical and illustrate a clear difference in the number of problems that students were able to complete. Students in the IT groups were able to work much more efficiently, or complete more problems in the same amount of time, than control group students.

This result suggests two different implications. The first is that completing more problems does not guarantee greater retention. Despite the fact that students in the IT groups were able to solve significantly more problems, they did not have a greater understanding of the material. This may suggest that the students' understanding of the problems was more shallow and that a shallow understanding of many problems is equivalent to a deeper understanding of a few problems. Alternately, perhaps the results imply that there is a limit to the number of problems that students can learn from and that once this limit is reached, any other problems are just busy work. A study designed to investigate repeated problem solving in the absence of the computer could potentially answer this question, but the results of the study described herein cannot.

Another implication of this finding on efficiency is that implementing IT allowed the instructors to expose the students to a number of different problems in a short amount of time without detrimentally affecting their learning and retention. This is an important finding because it not only highlights an important trait of IT, but it also suggests that IT can be used without hurting student performance. Thus, if an instructor would like to have students attempt more problems that they would ordinarily be able to solve by hand, using IT would allow them to solve a greater number of problems without hindering their understanding of the material. If instructional technology is implemented in a systematic manner, this research shows that IT use can be very beneficial in improving problem-solving efficiency.

For the following exercises, consider the truss in the figure below, which is support by a pin at H and a roller at D.



If θ = 45° and a 9kip force is applied vertically at *A* in the downward direction, determine the forces in GH, CH, and AF and indicate whether the member is in tension or compression.

GH = 4.50 kips ~~compression~~ tension

CH = 6.36 kips Compression

AF = 0 kips

What is the relationship between θ and the force in members CH and GH?

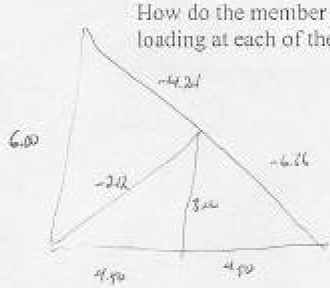
inverse: CH decreases in absolute value as θ increases

inverse: GH decreases in absolute value as θ increases

Do any of the member forces change if the 9kip force is applied at *F* rather than *A*; if so, how?

yes, AF goes under tension

How do the member forces change if the 9kip load is distributed across the bottom cord: 3kip loading at each of the joints *G*, *F*, and *E*?



AB decrease absolute value to -4.24 from -6.36, still under compress

AF increase absolute value to 6.00 from 0, now under tension

BD - no change

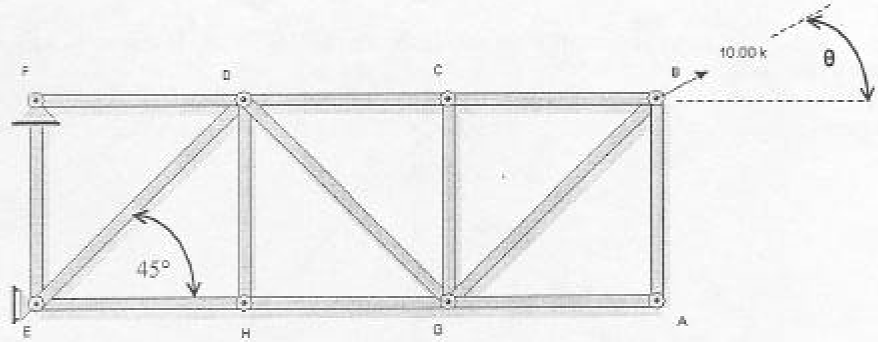
BE increase absolute value to 3.00 from 0, now under tension

BF increase absolute value to -2.12 from 0 now under compression

DE - no change

EF - no change

Figure 8.1 Example of Completed Special Session Exercises from Content-type Group



Consider the cantilever truss in the figure above which is supported by a pin at F and a roller at E and has an applied load of 10kips at B; if θ is equal to 30°, identify which members are in tension, which members are in compression, and which are zero force members.

tension : DE, EH, HG, DC, CB, GB

compression : EF, FD, DG

ZFM : DH, CG, GA, AB

Which members have the maximum compressive and tensile forces?

compressive : DG = - 7.070 kips

tensile : (tie) EH & HG = 10.00 kips

If θ is 0°, which members are in tension, in compression, or zero force members.

tension : FD, DC, CB

compression –

ZFM : everything else

If θ is 270°, which members are in tension, in compression, or zero force members.

tension : EF, FD, DG, DC, CB

compression : ED, EH, HG, GB

ZFM : DH, CG, GA, AB

If θ remains 270° but member BG is removed and replaced by a new member AC, how does this affect the force in AB? Does this change in configuration affect the maximum compressive and tensile forces in the truss?

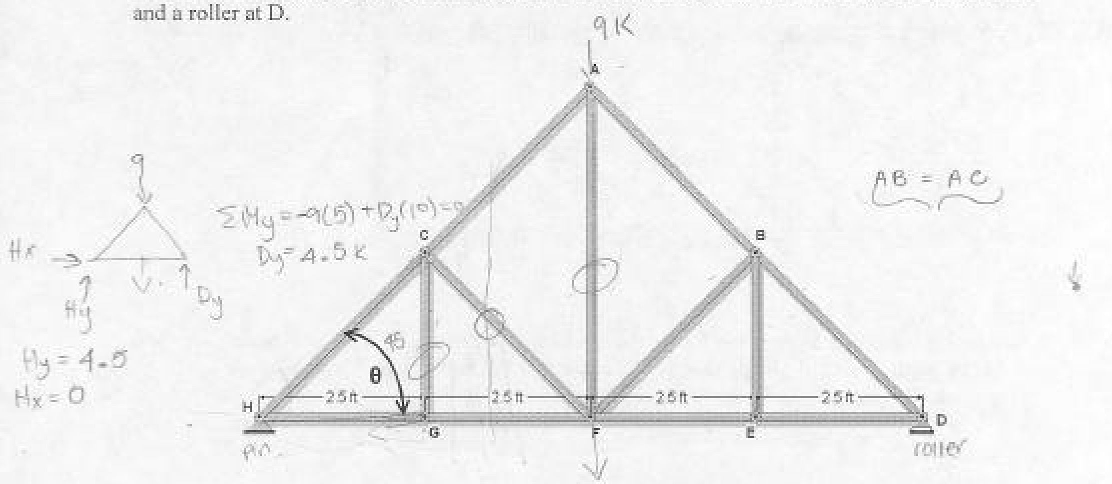
AB now has compression of 10 kips

compressive : FD = 30 kips

tensile : (tie) EH & HG = -20 kips

Figure 8.1 (cont.) Example of Completed Special Session Exercises from Content-type Group

For the following exercises, consider the truss in the figure below, which is support by a pin at H and a roller at D.



If θ = 45° and a 9kip force is applied vertically at *A* in the downward direction, determine the forces in GH, CH, and AF and indicate whether the member is in tension or compression.

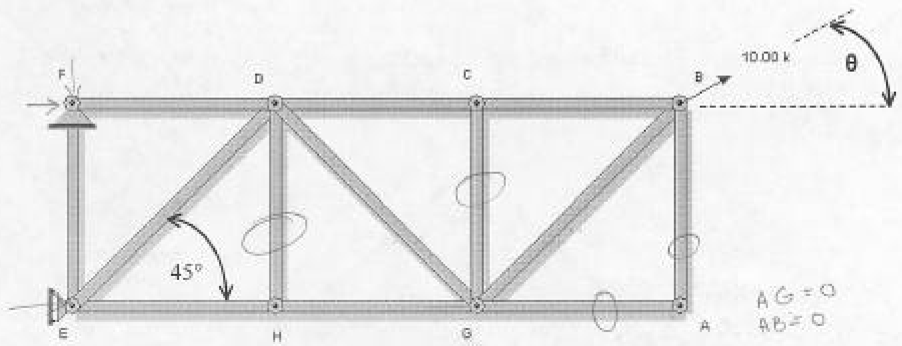
ΣFx = ACsin4 - ABsin45 = 0
AB = AC
-6.36 = 6.36
ΣFy = 0 = 9 - ACsin45 - ABsin45 - AF = 0
AF = 17.99K
ΣFx = 0 = -6.36 (cos45) + HG = 0
HG = 4.49
ΣFy = 4.5 + HCsin45 = 0
HC = -6.36K
HC = 6.36 (C)
HC = CA

What is the relationship between θ and the force in members CH and GH?

Do any of the member forces change if the 9kip force is applied at *F* rather than *A*; if so, how?

How do the member forces change if the 9kip load is distributed across the bottom cord: 3kip loading at each of the joints *G*, *F*, and *E*?

Figure 8.2 Example of Completed Special Session Exercises from the Control Group



Consider the cantilever truss in the figure above which is supported by a pin at *F* and a roller at *E* and has an applied load of 10kips at *B*; if θ is equal to 30°, identify which members are in tension, which members are in compression, and which are zero force members.

Which members have the maximum compressive and tensile forces?

If θ is 0°, which members are in tension, in compression, or zero force members.

If θ is 270°, which members are in tension, in compression, or zero force members.

If θ remains 270° but member BG is removed and replaced by a new member AC, how does this affect the force in AB? Does this change in configuration affect the maximum compressive and tensile forces in the truss?

Figure 8.2 (cont.) Example of Completed Special Session Exercises from the Control Group

8.7.3 Discussion of Findings across Time

The comparisons of the assessments across time also yielded unexpected results. As stated already, students performed at least as well on each subsequent assessment as they did on the previous one. More specifically, students performed better on the retention test than they did on the posttest. There was no difference, however, between scores on the retention test and long-term retention test. This suggests that students not only retained knowledge, but actually increased in the ten week interval between posttest and retention test. There are a couple of possible explanations for this, which are discussed below.

One possible reason for the increase in knowledge may be a test-retest bias. Such a bias may occur when students are asked to complete the same assessment multiple times. In some studies, it has been shown that students perform well on subsequent tests not because they remember the information but because they remember completing the exam. A test-retest bias may account for high retention rates, but is not assumed to account for an increase in performance. As such, it is assumed that a test-retest bias was not to blame, though a separate study would have to be completed to rule this possibility out completely.

Another possible reason for the increase is that something took place during the ten-week interval that helped students improve on the evaluations. No direct instruction on trusses occurred in any of the sections following the intervention, however, and so whatever may have influenced the students may have been indirect. There are two possible sources of indirect influence.

The first is the continued use of equilibrium principles. Though trusses were not directly addressed after the intervention, the solution methods used in truss analysis are based upon principles of equilibrium; these principles are the foundation of any course in statics and continue to be addressed in class even when truss analysis is complete. In each of the sections involved in the course, students continued to use equilibrium and thus may have been more comfortable applying these principles appropriately on the retention test.

The second, and most likely, source of additional knowledge was the midterm exams. Each of the study groups had a question, or two questions depending on the semester, on their midterm exams that questioned their truss analysis abilities. The midterm was completed between the posttest and retention test, was graded by the instructors and returned to the students with feedback. Whether the exam influenced student performance as a result of their studying for the test or as a result of their learning from the feedback is unknown and cannot be determined from this study. It is assumed, however, that the exam was responsible for the increase in performance as it was the only time during the ten-week or twenty-five-week interval when students used truss analysis and thus accounts for the increase on the retention test as well as the equality between the retention and long-term retention test.

8.7.4 Discussion of Findings on Question Types

Another unexpected result was that students from each treatment condition performed equally well on each of the problem types. It was theorized that students who used the software would be more comfortable completing qualitative problems as opposed to quantitative problems because they did a minimal amount of hand calculations during the special session. The opposite was also theorized, that the control students would perform better on the quantitative problems because these were the types of problems that they completed during the special session. These theories were rejected, however, because students from each group performed equally well on each problem type.

There are three possible explanations for this occurrence. The first is that the software did not enhance students' abilities to qualitatively analyze trusses as hoped. The second is that all three groups worked on the same problems during the special session and perhaps the problems themselves, rather than the media used to solve them, encouraged qualitative thinking, even in the control group. The third and final explanation is that students continued to use the quantitative skills they were accustomed to using in truss analysis to solve the qualitative questions. Though there was no way to determine definitively what caused this result, there was some anecdotal evidence to support the third explanation. The researcher observed while grading some of the assessments that some students would substitute values for the unknowns on the qualitative problems and then solve them in a quantitative manner. Further studies would have to be completed to determine which explanation is at the root of these findings; such a study would probably require interviews and/or focus groups to inquire into the students' thought processes while solving the different problem types.

The same conclusion can be drawn from these findings regardless of which of the three explanations is correct. That conclusion is that instructors should not rely on IT to help students think of problems qualitatively. Perhaps the best way to encourage students to think qualitatively is to introduce the trusses in a qualitative manner prior to teaching students how to solve them quantitatively. Whether such a method would be successful at encouraging students to think qualitatively about trusses cannot be determined from these results; a separate study would have to be conducted to answer that question.

8.7.5 Discussion of Survey Findings

Included in the retention test was an informal, non-scientific survey questioning students about their use of the software. No formal conclusions can be drawn from these results because the survey was not validated prior to use and thus could not be considered scientific. As stated in Chapter 7, a more formal survey was not conducted because the purpose of this research was to measure performance effects. Students' attitudes have been measured elsewhere and been found to be in favor of technology use (Sulbaran, 2002). The results from this study agreed with Sulbaran's results to some extent in that students seemed to enjoy using the software, thought it was somewhat beneficial, and somewhat recommended it for use in future classes. These results are positive and suggest that students had a positive experience with the software and enjoyed the break from lecturing and note-taking. Again, formal conclusions cannot be drawn from these results but can be found elsewhere (Sulbaran, 2002). More anecdotal evidence of the students' experiences are presented in Appendix G, which contains a complete listing of students' open-ended responses to a request for any comments they may have had about the software or trusses in general. The majority of these responses were positive and further support Sulbaran's formal conclusions (2002).

8.8 Research Contributions

The data, results, findings and conclusions that have been presented all point toward two main research contributions. The first contribution is a set of practical suggestions for instructors who are considering the implementation of instructional technology in the engineering classroom. The second contribution is the research process itself, which reveals how long-term retention studies can be conducted. These contributions are presented in greater detail in the remainder of this section.

8.8.1 Suggestions for Instructional Technology Implementation

A number of implications regarding future implementations of instructional technology resulted from this research. As stated at the beginning of this document, the purpose of the project was to provide specific reasons why technology should or should not be implemented in the classroom. The study was successful in that the findings did reveal specific reasons to implement technology and suggestions on how this can be done successfully.

The most important implication is that the use of technology does not detract from the learning experience. Students who used the software in class performed just as well on the assessments as students who did not use software, even 25 weeks after the intervention. This is beneficial because many instructors are hesitant to relinquish even a small amount of class time for such activities, but this research has shown that it can be done without any significant detriment.

Another implication is that software use allows students to work more efficiently. Thus, if an instructor wants to expose his or her students to a number of different problems, situations, scenarios, or case studies, IT appears to be an excellent way to do that. Combined with the fact that IT does not hinder performance, this fact becomes even more powerful; it suggests that students can be exposed to a greater amount of content, in a shorter amount of time, without affecting learning or retention.

Another practical suggestion for instructors stemming from this research is that an activity as simple as having students break into groups and practice problems in class can maintain high levels of retention. Again, some instructors have a problem doing this because it takes away from time that they wish to spend on lectures. This study, however, revealed that the benefits of conducting such activities appear to outweigh the costs. Activities as simple as in-class discussions, where the students talk as much as the professor, have been shown to improve learning by a number of different educational researchers (McKeachie, 1999, Biggs, 1999). While other fields of study have been accepting of such teaching methods, engineering instructors seem especially hesitant to relinquish control despite that fact that research conducted in engineering classes support research in other fields and reveal that in-class activities increase learning and retention (Baker et al., 1999). This research provides additional support to these findings and suggests that in-class activities should be used more often in engineering education.

In addition to these suggestions on what should be done, there are some implications on what should not be done as well. The first is that instructors should not implement IT for the sole purpose of increasing retention. Implementing IT, when done properly, is a very time consuming and a potentially expensive task; it takes time to review and select software, and once the software is selected more time and resources must be devoted to assuring that the students have access to it. The findings from this study, however, reveal that the learning and retention benefits of using IT are no different than those realized by simply conducting in-class activities, which is a much cheaper educational strategy that requires less development time. What can be concluded is that instructors should have a predetermined purpose for implementing software use other than increasing retention. One such purpose discussed previously is that IT can expose students to a greater variety of information in a shorter amount of time. Another purpose is that instructors may want students to familiarize themselves with a particular type of software that they may be required to use later in their careers. If instructors have a specific purpose for using software, this research suggests that such use is not a detriment. If instructors simply want to maintain high levels of learning and retention, however, simpler tasks such as employing in-class activities have proven to be just as effective.

Another similar finding revealed that IT should not be relied upon to help students think of problems in a qualitative manner. Though much more research is needed on this topic, this study suggested that using IT did not help students think of trusses qualitatively, as opposed to quantitatively, as instructors had hoped that it would. While the reason for this is unclear, the finding suggests that instructors may need to look elsewhere to help students gain a deeper understanding of trusses. It is also unclear whether this finding is transferable to other domains that are not as mathematically intensive as truss analysis. Because instructors are desirous of this type of deeper understanding, further studies should be conducted to determine why IT did not provide this deeper understanding and to determine how this type of knowledge can be acquired. Future research could also reveal whether this finding is valid in other fields of study as well.

The following list summarizes the practical implications of this research; it provides some important points that instructors should consider when deciding whether or not to implement instructional technology.

- Instructional technology should be implemented for a specific purpose (to allow students to solve many different problems in short amount of time for example)
- When implemented in a planned and systematic manner, instructional technology has been shown to maintain performance levels over time
- Instructional technology, however, should not be implemented solely to increase knowledge retention because there are easier ways to do this, as stated below
- Simple in-class activities have been shown to produce equally high rates of knowledge retention, (the chain of logic then circles back to the beginning, which suggests that IT should be used for a specific purpose)
- Instructional technology does not provide a deeper understanding of structural behavior; more research is needed in this area

The most important finding is that performance levels remained consistent over time, the related practical suggestion is as follows.

- Instructors should take measures to maintain high levels of retention, such measures include
    - In-class activities
    - Alternative instruction techniques
    - Instructional technology, when implemented systematically and appropriately

8.8.2 Suggestions for Long-term Retention Research

The second contribution that this project makes is that it exemplifies how long-term retention studies can be conducted in the classroom. As quoted previously, Semb and Ellis (1994) state that any new approach to teaching or alternative instruction strategy should be assessed to determine the effects on learning and long-term retention. While new instructional strategies are being implemented, most notably IT, few studies have been conducted to determine the long-term effects of these strategies. Perhaps this is because they are difficult and time-consuming to conduct, a notion that this research has affirmed. Or perhaps researchers are unsure as to how to control for the many non-experimental variables that are ever-present in classroom research. Perhaps educators simply do not know how to conduct a longitudinal study of knowledge retention.

This research has contributed to the field of engineering education by showing that long-term retention studies can be conducted. Furthermore, it also contributes by providing an example framework of how such long-term studies can be conducted in the classroom. Researchers can follow the process detailed in this document to conduct longitudinal assessments of knowledge retention effects resulting from any instructional strategy and in any field of study. While the process is explained in detail in earlier chapters of this document, there are a few practical suggestions that were of particular interest or concern in this work that are noted below.

- Assessment instruments must be assessed prior to being used formally to assure that they are usable, reliable, and valid. Exams or quizzes may be clear and concise to the instructor but may reveal some weaknesses when administered to students. It is better to discover these concerns in a formative stage of the project than when the actual data are being gathered.

- If the intervention involves courses taught by other instructors, it is important to get the instructors to agree to the research. To do this, they must understand the importance of the research and be willing to relinquish some control over their classes. They must also continue to be an informed and integral part of the research as their input and cooperation are vital to the success of the research.
- Longitudinal, classroom studies require extensive planning. A plan for controlling and/or accounting for non-experimental variables must be in place. A plan for tracking, contacting, and motivating students to participate in the long-term assessments must be in place. A plan for cooperating with instructors and maximizing the use of class time must be in place, etc.
- Students are reluctant to participate in assessments or evaluations on their own accord. Some form of external motivation must be provided to increase response rates of assessments. Following up with students in a subsequent course proved to be an effective and efficient way of gathering long-term assessments, but there were still a number of students who were not enrolled in the subsequent course. Proper planning and research are required to determine the most effective method of gathering long-term data.
- When multiple sections of a course are used, solicit the use of a single instructor to teach the topic around which the intervention is designed. Instructor biases are nearly impossible to account for in the data analysis and are potentially very damaging to the validity of the study.

- When conducting quasi-experimental research, replication adds a great degree of validity to the research. Replication serves to eliminate any biases resulting from non-random selection and assures that the results are not likely to be a result of non-experimental variables. Replication, however, is time consuming and thus must be appropriately planned for.

8.9 Suggestions for Future Research

As with any research project, the findings and conclusions resulting from this study suggest that there is more work to be done in the areas of knowledge retention in education and instructional technology. According to the literature reviewed in the course of this project, this study is one of the first to assess the effects of technology on retention at intervals of over one month long. This is also the only study found by the researcher to compare the use of a content-type software to the use of a tool-type software. Though the results revealed no difference between the effects of the two types of software used or between software use and traditional problem solving, the completion of another study similar to this one is merited. Mayer (2001) suggests the that results of no one single study should be relied upon and that multiple studies of similar design should be conducted in areas of IT and retention. Another study similar to the one described herein but conducted with students in a different field of study could certainly add validity to the findings reported in this document.

Furthermore, applying the framework used in this study to a different content domain could overcome some of the limitations of this study and lead to more firm conclusions about the long-term effects of IT. For instance, it was assumed that the control groups performed as well as the experimental groups because truss analysis is an active process that requires visualization, calculation, and a systematic thought process. If this study was replicated with a topic that was not so inherently engaging, the results would not only add to, or take away from, the validity of this research but more fully control for activity and thus further define the relationship between instructional technology and knowledge retention.

Another important area of research that stems from this study is the need to research long-term effects other than performance. This study assessed how students performed on posttests but it did not assess the thinking process that students went through while completing these instruments. It is possible that, as a result of the treatment condition, students may have taken a different approach to the posttest problems. Perhaps students who used the software were more like to use a particular analysis method than those who did not use the software. Such an evaluation would likely have to include interviews, focus groups, and talk-aloud protocols, all of which were outside the scope of this research. A more qualitative assessment of the internal effects of software use would greatly complement the quantitative performance measures discussed herein.

This research also supports Neisser and Hyman's (1999) call for more naturalistic research into long-term retention. While instructional technology has been proposed as a means of increasing long-term retention, this work and the work of others (Sulbaran, 2002) reveal that this is not always the case. More classroom-based, long-term research must be conducted in order to determine specific ways that retention can be improved in education.

More research into the differences between tool-type and content-type technology is required as well. Well designed and conducted studies could reveal the practical uses of each as well as the differing effects resulting from their use. Studies should be conducted to suggest the situations in which one type might be more appropriate than the other. This study attempted to reach such a conclusion with regards to retention but found that both types were equally effective. There may, however, be other variables or situations that may suggest the implementation of one type as opposed to the other. Such situations and suggestions can only stem from future research projects.

More specific areas of future research developed from this project as well. One area of interest is the degree to which exams affect the learning process. It was assumed that the increase in scores from the posttest to the retention test were a result of the midterm exams that each of the students took. This assumption, however, could not be tested with the data gathered in this project. A future study could be conducted in a similar manner, but collect retention data shortly after midterm exam, thus shortening the retention interval and centering the interval around the midterm exam. Doing this could lend more confidence to the assumption that examinations resulted in better performance. Such a result would lead to the conclusion that exams are not only assessment instruments, but are learning experiences as well—research in this area may have been done before but as this was not pertinent to the development of this study, a literature review in this area was not completed. Furthermore, controls could be put in place to determine if the increase was a result of students studying for the exam or learning from the feedback that the students received from the instructor about the exam.

Another area worthy of future investigation is one that statics instructors perceived as important, the qualitative understanding of the structural behavior of trusses. As discussed in previous chapters, it was assumed that instructional technology could help students view the truss as a whole rather than as a construction of separate joints or sections and that this understanding would help students determine conceptually how the entire truss would behave in certain situations. Unfortunately, the study revealed that this goal was not accomplished and that IT has no effect on the students' ability to qualitatively analyze trusses. More extensive and more personal research, such as interviews and focus groups, may reveal the students' though processes and determine why this hypothesis did not hold true. Additional studies could also be conducted to determine other ways to help students learn qualitative analysis. The long-term effects of other methods, such as working with actual trusses or looking at trusses qualitatively in class before discussing them quantitatively, could be investigated by adjusting the framework of this study to meet such goals and objectives.

One final area of proposed future research is on the topic of practice. This research showed that students who solve many problems on the computer performed no better on performance and retention assessments than students who completed a few problems by hand. It is unknown whether the equality of results was due to a shallower understanding of the experimental students or if completing additional problems beyond a certain point did not add to a student's understanding of the analysis. One way to converge on the cause would be to conduct a study that rules out instructional technology as a variable and compares students who complete many problems by hand to students who complete a few problems by hand. The number of problems could be varied by groups of students and an optimal amount of practice could be determined, though this optimal number may vary by content domain. This would have very practical implications on the assignment of homework and could be very beneficial.

In conclusion, there is much more research that needs to be done. The link between instructional technology and knowledge retention remains ill-defined and only future studies conducted in the classroom can further describe the relationship. Furthermore, the nature of long-term retention and the factors that affect it remain in debate. The researcher joins Neisser (1999) and Semb and Ellis (1994) in the ongoing request for more practical studies in the area of long-term retention.

CHAPTER 9

SUMMARY AND CONCLUSIONS

This study was designed to investigate a possible link between instructional technology use and long-term knowledge retention. Literature in engineering education suggested that IT use improved learning in some cases and further theorized, by extension, that IT could improve retention of knowledge as well. There were, however, little or no studies to support this theory. This study was designed to test the theorized relationship and fill an important gap in the engineering education literature.

The study was also designed to investigate whether the use of a tool-type IT would result in different degrees of learning and retention than a content-type IT. Tool-type software are programs that are designed to complete a specific, non-educational task. Structural analysis software and spreadsheet programs are examples of tool-type IT. Content-type IT are programs that are developed specifically for educational purposes and do not have a *real world* task. Electronic textbooks and intelligent tutors are examples of content-type IT.

Based on evidence from studies in psychology, it was determined that the research should be conducted in a naturalistic setting so as to obtain a greater degree of validity. The experiment was designed to take place in the classroom with entire sections of students serving as sample groups. As such, the experiment was quasi-experimental and was replicated in subsequent semesters to overcome possible selection biases. The experiment involved implementing IT in a class setting and then testing students at 10 and 25 weeks to see if there were any differential effects in learning and retention as a result of the IT use. These retention intervals were chosen because the end of the intervals occurred at the end of the semester and the middle of the next semester respectively; times at which students are expected to remember information taught in any course.

The class that was chosen as the setting for the experiment was a statics and dynamics class. This class was chosen for several reasons. There were many sections offered each semester at Georgia Tech, the class was a low level course meaning that students would still be on campus for the long-term assessments, and because the course was so ubiquitous. To further contain any non-experimental effects and to simplify assessments, one topic within the statics curriculum was chosen as the focus of the study. Truss analysis was chosen because the topic is easy to assess, is often difficult for students, and is a topic that is included in most commercially available software packages designed for use in statics courses.

The research was designed to measure four variables. The first variable was prior knowledge and it was assessed via a pretest administered immediately before the intervention. Learning was measured via a posttest immediately after the intervention. Retention was tested via a test identical to the posttest ten weeks after the intervention. Long-term retention was assessed via a test identical to the posttest and retention test 25 weeks after the intervention. The assessment instruments were tested for usability, reliability and validity in a formative assessment that took place during a summer term prior to the first administration of the actual study. Prior learning was measured as a means of equating each of the groups; if each of the groups performed equally well on the pretest then it could be concluded that, despite the lack of random sampling, the groups were equivalent.

It was hypothesized, based on the literature, that IT use would increase posttest scores. By extension, it was also hypothesized that IT use would also increase retention and long-term retention test scores. It was also hypothesized that the content-type group would perform equally as well as the tool-type group on all of the assessments.

The research design included an intervention that would take place in three study groups. One group used content-type software, which is software developed specifically for use in education. Another used tool-type software, which is software that was not developed for education but was commercially developed to complete a “real world” task, such as structural analysis. The third group used no software during the special session of the intervention. The complete intervention took place during the period of time in which trusses are normally taught, or four lecture hours. During two of these hours, standardized lectures about trusses and truss analysis were presented to each of the three study groups. The same lecture was provided to each of the three sections and all lectures were given by a guest lecturer to avoid any bias resulting from differing instructors. The third session of the intervention was the special session, the session in which the software would be introduced to the experimental groups. A group of problems was developed for use in the special session and students completed the problems differently depending upon the group they belonged to. The control group worked on the special session problems in class, in groups of two or more, and used hand calculations. The content-type group completed the special session exercises in the computer lab, in groups of two or more, and with the assistance of the content-type software. The tool-type group completed in the special session exercise in the computer lab, in groups of two or more, and with the assistance of the tool-type software. Students in the experimental groups were able to complete more the of special session exercises than students in the control groups.

The software was chosen very carefully and with the assistance of six statics instructors. The statics instructors identified some key functionalities that they felt were essential elements of a software package about trusses. Some important elements of instructional software as dictated by educational theory were also identified. These functionalities and elements were compiled into a list that was then used to select two software titles, one content-type and one tool-type, from a number of different readily available programs.

Following the special session, the last session of the intervention was used to collect the posttest, which students completed in class. The class was then returned back to the regular instructors who continued to teach the course without referring to trusses in class except in reference to midterm exams. In each of the three sections, during each administration of the experiment, identical truss analysis questions were included on midterm examinations, which were graded and returned to the students with feedback as with most midterm exams. Trusses were not addressed in class during the remainder of the semester. Just prior to the end of the semester, the retention tests were given to the students who completed them at home in a limited amount of time and returned them within a few days.

The long-term retention test was administered 25 weeks after the intervention. At this time, students had begun a new semester and were enrolled in different classes. The long-term retention test was administered to civil and environmental engineering students in a follow up course. Students took the long-term retention test home and completed it in a limited amount of time and returned the exam within a few days. Students who were not enrolled for the follow-up course in which the test was administered, were emailed a copy of the long-term retention exam and asked to complete and return it at their earliest convenience.

The study was replicated, or conducted again, in the subsequent semester. Replication was necessary to overcome the inherent weakness of the quasi-experimental design. The replicated experiment was identical initial implementation except, of course, that it was conducted with different student groups.

The data gathered during the four assessments, and the exam questions as well, were all analyzed according to accepted parametric and nonparametric statistical procedures. Pretest scores were compared across study groups in each of the two studies and no significant differences were found, suggesting that each of the groups possessed equal amounts of prior knowledge. In other words, each of the groups began the truss analysis portion of statics with the same knowledge base. Other factors including major, student rank, gender, ethnicity, credit hours, and GPA were compared to assure equality of groups. Most of the variables were equal among all groups with two exceptions. In the fall study, the groups differed significantly in terms of rank and in the spring study the groups differed significantly in terms of major. It was assumed that replication would account for these differences or that if results were consistent for both studies then these differences could not be the cause because they were not manifested in both studies. The results were consistent in both studies and so no further investigation into the differences of rank and major was conducted and the groups were assumed to be equal.

Posttests, retention test, and long-term retention test scores were also compared across treatment groups. In both implementations of the study, there were no significant difference between the scores on any of the assessments. Each of the groups performed equally well on each of the evaluations. Three types of questions appeared on these tests: qualitative, quantitative, and factual. The scores on each of the question types were compared across treatment conditions as well with similar results: no significant differences. Differential scores for the retention and long-term retention tests, the difference in scores between these respective tests and the posttest, were also compared across treatment conditions and no significant differences were found. Unexpectedly, the mean differential scores were positive which meant that students performed better on the retention tests than they did on the posttest. This prompted a comparison of scores across time, revealing that retention test scores were significantly greater than posttest scores but were not significantly different from the long-term retention test scores.

These results support the following findings and conclusions:

- Students who used IT did *not* approach problems in a manner that was qualitatively different than students who did not use IT
- Students who used IT had high rates of retention
- Students in the control groups who spent class time solving problems in groups by hand had equally high rates of retention
- Using in-class activities was just as effective at maintaining retention and long-term retention as using IT

- IT promoted efficient problem solving, allowing students to solve more problems in a shorter amount of time, without hindering learning or retention

In conclusion, this research has a few important practical recommendations for engineering educators. Retention of course knowledge can be achieved by implementing an activity as simple as having students pair off and solve problems during class. Instructional technology can also help students retain knowledge but not by a greater amount than in-class activities. Because instructional technology requires more planning and resources than other in-class activities, it should be implemented for a specific purpose other than to increase retention. One such purpose that was revealed in this study is that software use allowed students to solve more and different problems in a short amount of time. If an instructor does choose to implement instructional technology, this study suggests that it will not hinder the educational experience and students will retain much of what is learned.

APPENDIX A

SOFTWARE EVALUATION FORM

NOTE: This form was developed and tested, but not used because the tests proved it to be unreliable. Please refer to Chapter 4 for more details.

Interface

	Agree		Neutral		Disagree
The software interface is easy to use and requires little cognitive demand. *The system should allow the user to devote much of their cognitive resources to accomplishing the task at hand rather than understanding the interface.*	1	2	3	4	5
Navigation through the program is intuitive. *The user should be able to move throughout the system without getting lost or confused.*	1	2	3	4	5
It is clear what the software is doing and what is expected from the user. *The user should understand the status of the system and how that status relates to the task.*	1	2	3	4	5
The program uses consistent and accepted standards. *The system should use standard buttons, icons and menus whenever possible.*	1	2	3	4	5
The software can be used without an excessive amount of instructions. *Is the amount of learning that is required to use the software appropriate for the task?*	1	2	3	4	5

Comments:

Innovation

	Agree		Neutral		Disagree
The program is highly interactive. *A highly interactive system is one in which the user controls the environment or can control program parameters as opposed to a minimally interactive system such as a slide show where the user simply clicks a button to receive the next piece of information.*	1	2	3	4	5
The program utilizes multiple representations for learning. *Multimedia representations should be utilized to describe information, thus helping students with different learning styles.*	1	2	3	4	5
The program utilizes real-world representations and visualizations. *Pictures, simulations, and diagrams should be authentic.*	1	2	3	4	5
The program utilizes up-to-date software technology. *The system should resemble other educational and application software that the user is accustomed to using and multimedia objects should use the latest technologies.*	1	2	3	4	5
The program allows the user to explore and experiment. *The software should give users opportunities to design, construct, or manipulate objects.*	1	2	3	4	5

Comments:

Introspection

	Agree		Neutral		Disagree
The user is allowed to control the pace of the interaction. *The user should be in control at all times.*	1	2	3	4	5
The software provides helpful feedback, which allows user to recognize and correct mistakes. *Feedback should consist of more than just an error code and should be constructive rather than simply correcting user mistakes.*	1	2	3	4	5
The program requires the user perform tasks that involve thinking and reflection. *The user should demonstrate mastery by solving problems, answering questions or accomplishing a particular task.*	1	2	3	4	5
The task required of the user is authentic and cognitively appropriate. *The task should resemble a 'real world' task and should be of appropriate difficulty.*	1	2	3	4	5
The software includes support features to help user accomplish the desired task. *Constructive support should be provided at critical points or transitions to help user continue to use the system with ease.*	1	2	3	4	5

Comments:

Integration

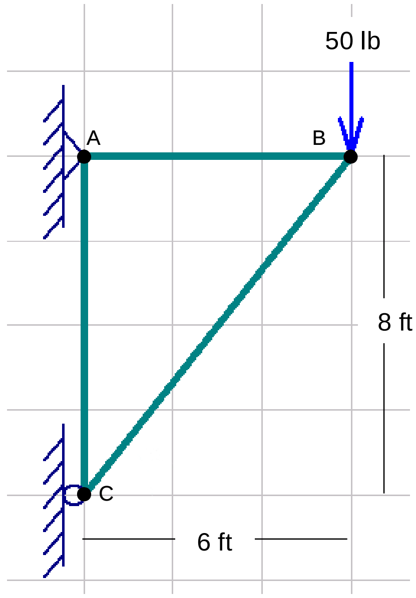
	Agree		Neutral		Disagree
The instructor is allowed to customize the program. *The program or presentation should fit the instructor's course rather than the course fitting the program.*	1	2	3	4	5
The program content integrates well with the course content. *The skills, information or processes taught or used by the program should be identical to those taught in the course.*	1	2	3	4	5
The sections of the program integrate well with other sections of the program. *The program should tie different sections together rather than representing them as distinct units.*	1	2	3	4	5
The program integrates well with the curriculum. *References to issues outside of the course should be authentic and not contradict information that will be taught in future courses.*	1	2	3	4	5
The program integrates well with the profession. *Information should be consistent with accepted ideas and practices.*	1	2	3	4	5

Comments:

APPENDIX B

PRETEST IN PRELIMINARY FORM

Find the forces in members *AB* and *BC*. Show all work.



Members of a truss are assumed to be joined by smooth pins.

_____ True _____ False

Truss loadings are assumed to be applied at the joints only.

_____ True _____ False

All truss members are assumed to be two force members.

_____ True _____ False

List the equations of equilibrium that can be used when analyzing a joint in a truss.

Force is a scalar quantity?

_____ True _____ False

A vector has both a magnitude and a direction.

_____ True _____ False

Any vector can be broken down into Cartesian components.

_____ True _____ False

Vectors can be added:

_____ by using the parallelogram law.

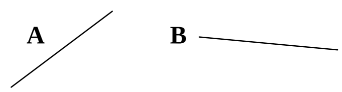
_____ by using a triangle construction

_____ by summing the *x* and *y* components of the vectors.

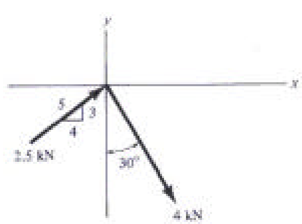
_____ all of the above

Which of the following is the resultant of vectors **A** and **B**?

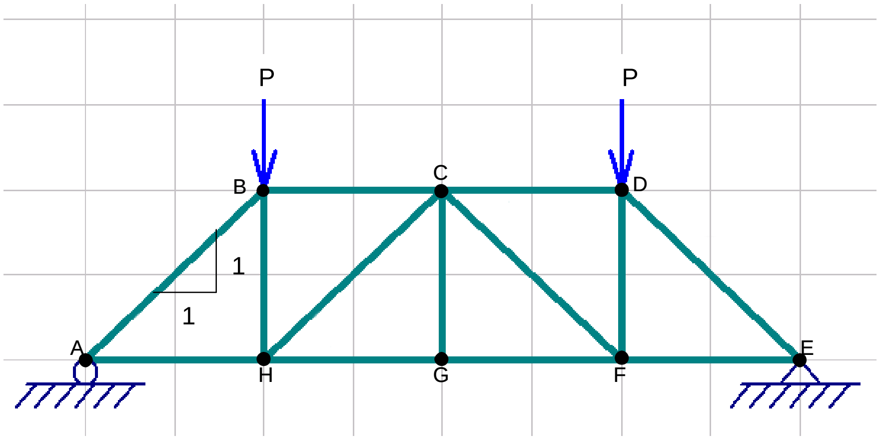
(Circle one)



Resolve the following force vectors into their x and y components. Add the vectors and determine the magnitude and direction, Q measured from the positive x axis, of the resultant force.



Assuming that the forces acting on the truss below are in the direction shown and are equal, answer the following questions.



Member *CG* is:

_____ In Compression

_____ In Tension

_____ Zero Force Members

The members that make up the lower chord of the truss (*AH*, *EF*, *FG*, & *GH*) are:

_____ In Compression

_____ In Tension

_____ Zero Force Members

Member *CF is*:

_____ In Compression

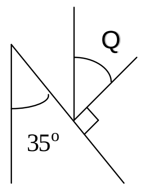
_____ In Tension

_____ A Zero Force Member

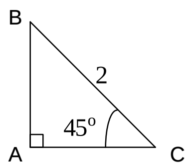
Solve the following equations for x and y.

$$2x + 3y = 12 \qquad\qquad x + y = 3$$

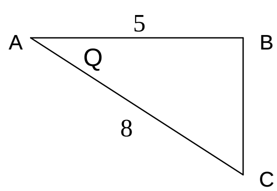
What is the value of Q in the figure below?



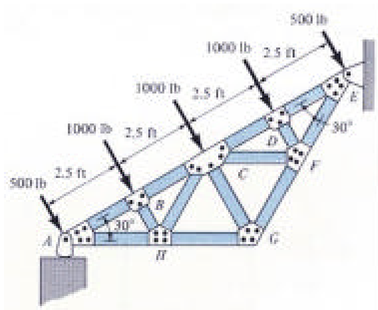
What is the value of AB in the figure below?



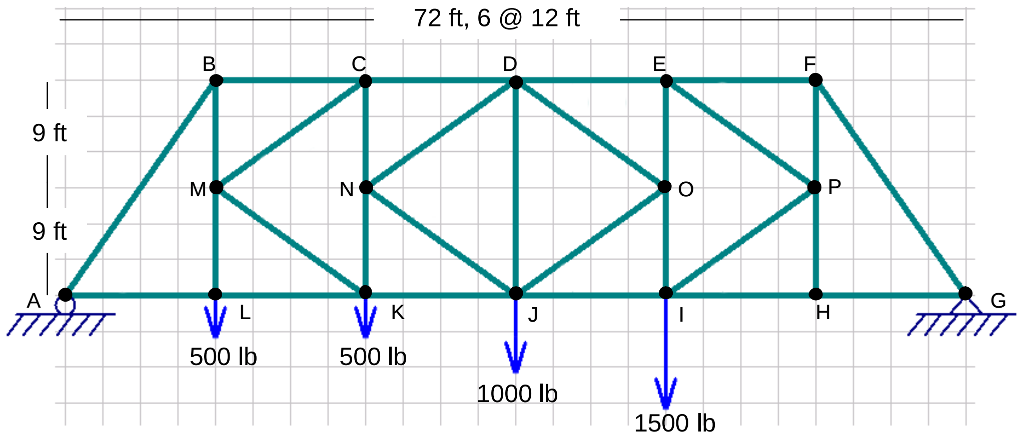
What is the value of Q in the figure below?



The structure below is in equilibrium. Solve for the reactions at both supports.



Determine the forces in members *ML*, *HP*, *EF* & *HI* and whether they are in tension or compression. Show all work.

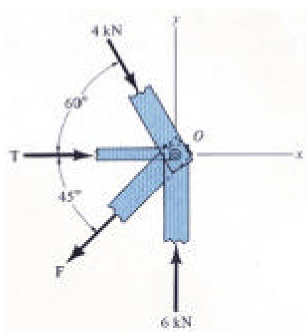


List the independent equations of equilibrium that are used in 2D problems.

How many reaction forces do the following types of supports provide?

Roller	__________
Pin	__________
Rocker	__________
Fixed	__________

The members of the truss are pin-connected at joint *O* as shown below. Determine the magnitudes of F and T to satisfy equilibrium.



APPENDIX C

FINAL FORM OF POSTTEST

PLEASE READ THESE INSTRUCTIONS IN THEIR ENTIRETY AND SIGN YOUR NAME TO VERIFY THAT YOU HAVE DONE SO.

This assignment is part of a research project being conducted in the School of Civil and Environmental Engineering at Georgia Tech. The purpose of this study is to understand and improve the way teaching and instruction is done in the school. As each of you are CEE students, this study could potentially have a direct impact on your education but only if you participate.

Your participation is essential to the completion of this study. The results of this study will provide insights to the School of CEE, the College of Engineering, the Associate Provost for Institutional Development and others who are responsible for making important decisions regarding instructional, curricular, and program development.

Further, your completion of this assignment will count towards your CEE 3020 class participation requirement, which is 5% of your final grade.

Please take **only 40 minutes** to complete the attached assignment, this is probably not enough time to complete all the problems but please do as much as you can in that amount of time. Please use only a calculator when completing this assignment (no books, notes, or help from other persons). Please give each of the problems on this assignment your honest effort so that an accurate assessment can be made.

As always, your confidentiality is assured. Aggregate data only, with all names removed, will be published to offices on campus or off.

You must complete and return the assignment by September 19, 2003 in order to get credit. Return the assignment to Sean St.Clair's mailbox, on the third floor of the Mason building at the end of the hall, near Dr. Kurtis' office. If you have any questions, please send them via email to sean.stclair@ce.gatech.edu. Thank you for your support and participation.

__________________________ __________________________

Signature Date

____

Printed Name

Problem 1. Answer the following questions about trusses in general.

Members of a truss are assumed to be joined by smooth pins.

_____ True _____ False

Truss loadings are assumed to be applied at the joints only.

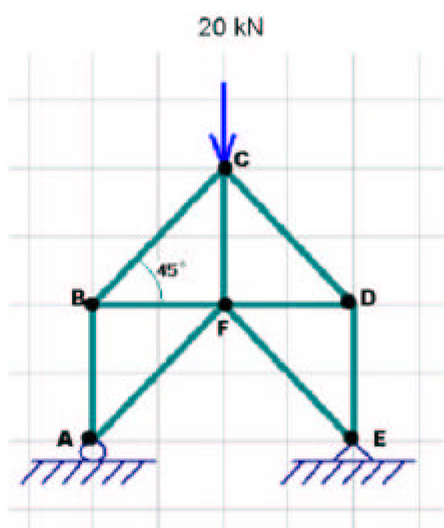
_____ True _____ False

All truss members are assumed to be two force members.

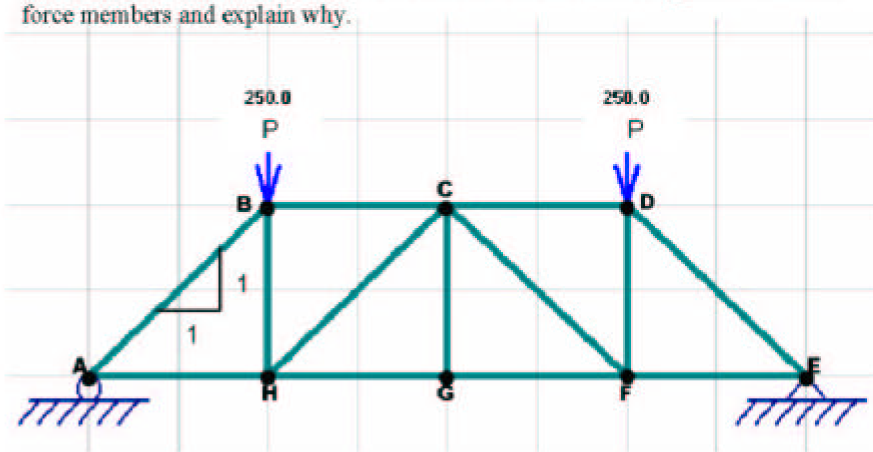
_____ True _____ False

List the equations of equilibrium that can be used when analyzing a joint in a truss.

Problem 2. Determine the member forces in the truss below and indicate whether they are in tension or compression. Diagram is drawn to scale, all non-right angles are 45°. Show all work.



Problem 3. For the following truss which is supported and loaded as shown, determine whether the following members are in tension, in compression or zero force members and explain why.



Member *CG* is: Explain why:

_____ In Compression

_____ In Tension

_____ Zero Force Members

The members that make up the lower chord of the truss (*AH*, *EF*, *FG*, & *GH*) are: Explain why:

_____ In Compression

_____ In Tension

_____ Zero Force Members

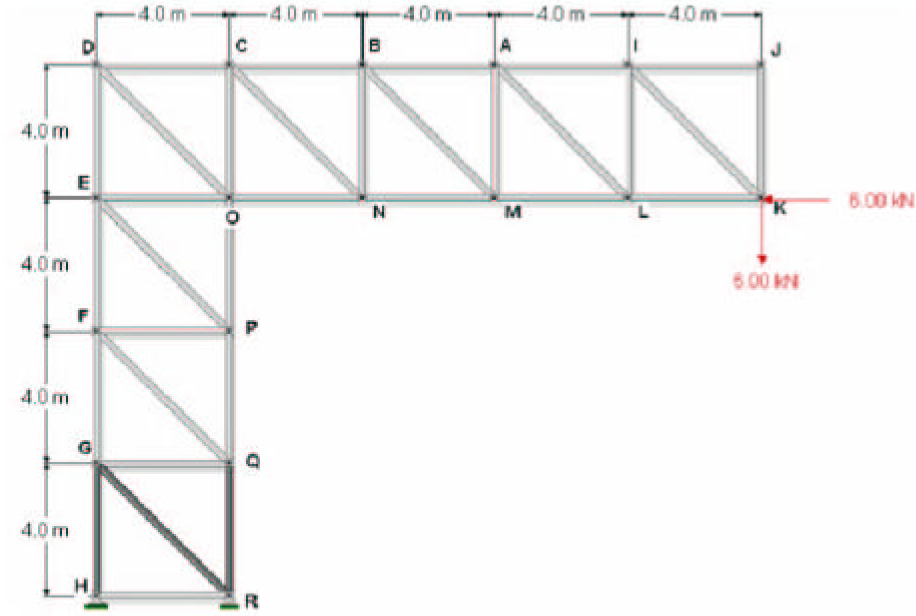
Member *CF is*: Explain why:

_____ In Compression

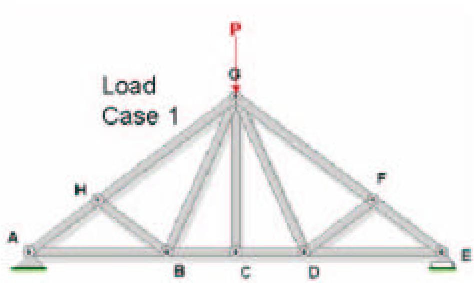
_____ In Tension

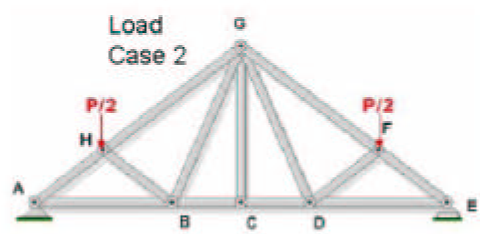
_____ A Zero Force Member

Problem 4. The truss below is supported by a pin at H and a roller at R (simply supported). Determine the forces in the highlighted members *GH*, *GR* & *QR*. Indicate whether the members are in tension or compression. Show all work.



Problem 5. Consider the simply supported fink truss below that may be subjected to two different load cases.





For the members listed below, discuss the differences in the internal forces due to the different loading conditions and why (i.e. under load case 1, member *XZ* is in tension, under load case 2 it is in compression because…, or the magnitude of XZ is twice as big in case 2 as it is in case 1 because…).

Member *BG*

Member *BH*

Member *CG*

Member *GH*

Member *AB*

APPENDIX D

FINAL FORM OF PRETEST

PLEASE READ THESE INSTRUCTIONS IN THEIR ENTIRETY AND SIGN YOUR NAME TO VERIFY THAT YOU HAVE DONE SO.

In order to improve instructional methods in the College of Engineering and thus make Georgia Tech engineers even more knowledgeable and competitive, a research project is being conducted in a number of different sections of CEE2020.

As engineering education continues to evolve and change, it is important for educators to understand the nature of how students learn and use this understanding in our instructional methods. The results of this and other assignments will provide insights to the College of Engineering, the Office of Assessment, the Associate Provost for Institutional Development and others who are responsible for making important decisions regarding instructional, curriculum, and program development.

Your participation is a required part of the course and is worth 5% of your final grade. To earn these points, you must complete three assignments, the first of which is attached as a take-home assignment. The second assignment will be done in class and the third will be given as a take-home assignment later in the semester. This assignment should take about 25-40 minutes to complete. Your honest effort to answer each question correctly will provide us with accurate results that will be used to improve the educational experiences of students at Georgia Tech.

Your instructor will be the only person who will have access to your individual results. You may be assured of complete confidentiality; aggregate data only, with the names removed, will be published to offices on campus or off.

This assignment is to be performed by you without the help, in any fashion, of any other person. The use of your textbook, course notes, or any resource other than a calculator is not permitted. By signing your name below you acknowledge that you have read and agree to the above conditions.

____________________________ ____________________________

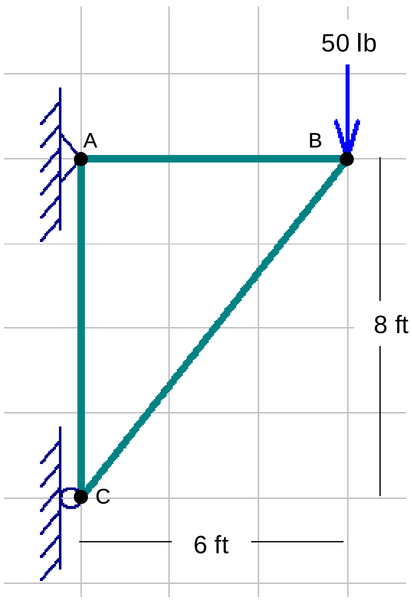
Signature Date

____

Printed Name

State the time at which you started the assignment ____________

Find the forces in members *AB* and *BC*. Show all work.



Force is a scalar quantity?

______ True ______ False

A vector has both a magnitude and a direction.

______ True ______ False

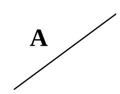
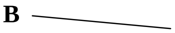
Any vector can be broken down into Cartesian components.

______ True ______ False

Vectors can be added:

- ______ by using the parallelogram law.
- ______ by using a triangle construction
- ______ by summing the *x* and *y* components of the vectors.
- ______ all of the above

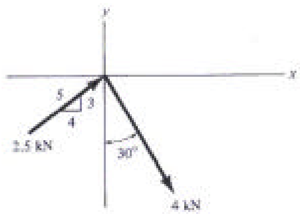
Which of the following is the resultant of vectors **A** and **B**?
(Circle one)

A  **B** 

If $\mathbf{r} = 5\mathbf{i} + 3\mathbf{j} + 5\mathbf{k}$ and $\mathbf{F} = 38.2\mathbf{i} + 420\mathbf{k}$, then $\mathbf{r} \times \mathbf{F} =$

For the following force system…

1. Resolve the forces into their x and y components.



2. What is the vector form of the resultant force?

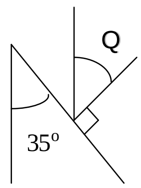
3. Calculate the magnitude of the resultant force.

3. Determine the direction (Q measured from the positive x axis) of the resultant,

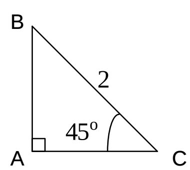
Solve the following equations for x and y.

$$2x + 3y = 12 \qquad\qquad x + y = 3$$

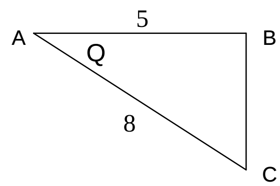
What is the value of Q in the figure below?



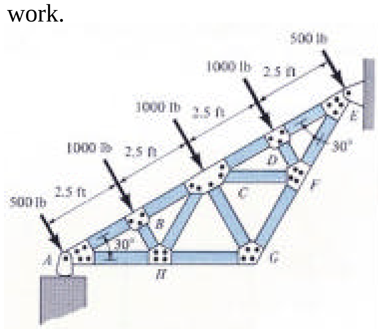
What is the value of AB in the figure below?



What is the value of Q in the figure below?



The structure below is in equilibrium. Solve for the reactions at both supports. Show all work.

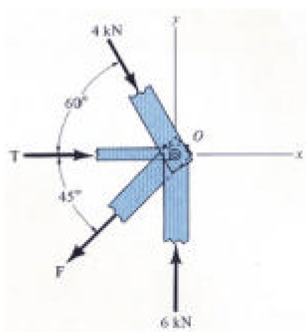


List the independent equations of equilibrium that are used in 2D problems.

How many unknowns are associated with the following types of supports?

Roller	__________
Pin	__________
Rocker	__________
Fixed	__________

The members of the truss are pin-connected at joint *O* as shown below. Determine the magnitudes of F and T to satisfy equilibrium. Show all work.



State the time at which you completed the assignment ______________

APPENDIX E

RETENTION TEST COVER LETTER AND DEMOGRAPHIC QUESTIONNAIRE

PLEASE READ THESE INSTRUCTIONS IN THEIR ENTIRETY AND SIGN YOUR NAME TO VERIFY THAT YOU HAVE DONE SO.

This assignment has a time **limit of 40 minutes**. Please time yourself starting **after** you complete the demographic information questions on the next page.

This is the third and final assignment that constitutes your participation in a research project being conducted in various sections of CEE2020. Your participation is a required part of this course and is worth 5% of your final grade. To earn these points you much complete this assignment. Your honest effort to answer each question correctly will provide us with accurate results that will be used to improve the educational experiences of students at Georgia Tech.

Your instructor will be the only person who will have access to your individual results. You may be assured of complete confidentiality; aggregate data only, with the names removed, will be published to offices on campus or off.

This assignment is to be performed by you without the help, in any fashion, of any other person. The use of your textbook, course notes, or any resource other than a calculator is not permitted. By signing your name below you acknowledge that you have read and agree to the above conditions.

__________________________ __________________________

Signature Date

____

Printed Name

What is your grade/rank?

☐ Freshman ☐ Sophomore ☐ Junior ☐ Senior

What is your GPA?

☐ 3.5-4.000 ☐ 3.0-3.499 ☐ 2.5-2.999 ☐ 2.0-2.499 ☐ 1.5-1.999 ☐ 1.0-1.499

How many credits are you taking this term? ______________________

What is your ethnicity?

☐ Asian/Pacific Islander ☐ African American ☐ Caucasian
☐ American Indian ☐ Hispanic
☐ Other: ______________________

Did you seek outside help for the truss portion of this class?

☐ Yes ☐ No

If so, from where did you receive help?

☐ Help session ☐ Private tutoring
☐ One-on-one help from instructor ☐ Classmates
☐ Other: ______________________

How many times did you receive help during the truss portion of the class? ________

Did you seek outside help for the other portions of this class?

☐ Yes ☐ No

If so, from where did you receive help?

☐ Help session ☐ Private tutoring
☐ One-on-one help from instructor ☐ Classmates
☐ Other: ______________________

How often did you seek outside help for this class?

☐ More than once a week ☐ A couple of times a month
☐ Once a week ☐ Once or twice during the semester

	No		Uncertain		Yes
Did you enjoy using the software?	1	2	3	4	5
Did you use the software for topics other than trusses?	1	2	3	4	5
Would you use the software again?	1	2	3	4	5
Are you still using the software?	1	2	3	4	5
Would you recommend using the software in future classes?	1	2	3	4	5
Do you think the software helped you understand trusses better?	1	2	3	4	5

Please include any general comments you have about trusses, the software you used, or your experience with this class (use back if necessary).

____

APPENDIX F

LONG-TERM RETENTION TEST COVER LETTER

PLEASE READ THESE INSTRUCTIONS IN THEIR ENTIRETY AND SIGN YOUR NAME TO VERIFY THAT YOU HAVE DONE SO.

This assignment is part of a research project being conducted in the School of Civil and Environmental Engineering at Georgia Tech. The purpose of this study is to understand and improve the way teaching and instruction is done in the school. As each of you are CEE students, this study could potentially have a direct impact on your education but only if you participate.

Your participation is essential to the completion of this study. The results of this study will provide insights to the School of CEE, the College of Engineering, the Associate Provost for Institutional Development and others who are responsible for making important decisions regarding instructional, curricular, and program development.

Further, your completion of this assignment will count towards your CEE 3020 class participation requirement, which is 5% of your final grade.

Please take **only 40 minutes** to complete the attached assignment, this is probably not enough time to complete all the problems but please do as much as you can in that amount of time. Please use only a calculator when completing this assignment (no books, notes, or help from other persons). Please give each of the problems on this assignment your honest effort so that an accurate assessment can be made.

As always, your confidentiality is assured. Aggregate data only, with all names removed, will be published to offices on campus or off.

You must complete and return the assignment by September 19, 2003 in order to get credit. Return the assignment to Sean St.Clair's mailbox, on the third floor of the Mason building at the end of the hall, near Dr. Kurtis' office. If you have any questions, please send them via email to sean.stclair@ce.gatech.edu. Thank you for your support and participation.

________________________ ________________________

Signature Date

____

Printed Name

APPENDIX G

STUDENT COMMENTS AND SUGGESTIONS

The following were responses to the following statement on the questionnaire:

“Please include any comments you have about trusses, the software you sued, or your experience with the class.”

- was not useful for class
- it was hard to understand at first
- portion on trusses was taught
- more helpful if I could use it on my own computer
- I didn't really understand trusses because too little time was spent on it
- software was great
- only one time with software
- helpful, should be used more than once
- helped visualizing
- software helped better understand tension and compression when forces were applied to different members
- more user friendly software would have been better, and surely better if presented earlier in the semester
- the software helps to picture the forces acting on a truss in real time.
- helpful for the overall picture, but not in the hw because we had to do our own calculations anyway
- this class doesn't excite me (hence I am an IE major) which is probably why my grades aren't so good
- I needed more time to look at what I was doing as well as the problem itself.
- the software helped me visualize what was happening
- visualization of the computer helped out a lot.
- I did worse because of the change of the instructor
- I don't think there was enough time to study what we were drawing and see how it was effected by adding or subtracting members
- I think it was helpful in trusses section
- software helps you train your intuition more quickly
- should have 2 labs instead of one
- made problems easier to solve… not understand better (like a calculator)
- it would have been more useful if I used it outside of class, but I didn't
- I can't tell which are in tension, which are in compression…
- I cannot stand trusses!
- I think if I could review my notes for 20 mins, I would completely understand it
- I enjoyed the truss portion of the class, thought it was taught well. Although, my retention isn't great
- I don't know much at all
- I can't really remember that well without looking at the book
- only used it one time so it didn't help me
- the software was a little confusing, but we only did it one day, and it was a long time ago, so it's hard to give it a good assessment

- I wasn't in class the day they used the software
- the software explained trusses well and gave us an opportunity to check out work.
- the software was very helpful because it gives you an interactive and visual approach to trusses. I would recommend using it on other sections, maybe have one class every week or two devoted to a "lab" where the software is used. I have enjoyed this class and learned a lot. I also really enjoyed the teacher and his way of teaching
- I think the software would be helpful however with the bugs in the program it probably would prove to more of a headache than a help
- Trusses are very interesting, but difficult to grasp at different levels of applications. The software application help me by allowing me to see immediate result with manipulation. This help to better understand typical behaviors of truss systems.
- In my experience there were still a few bugs to work out but it could be an extremely useful tool for students to check their work against.
- give examples of how to use it for other aspects of the class
- I thought that trusses was the easiest part of the course. I thought that the software was good. It was fun to use and helped.
- the software was great. Wish I had used it more.
- I think the software may be helpful in a practical sense outside of the classroom.
- I enjoyed the section about trusses the most in this class. I think a little more instruction was needed before using the software
- I would have liked to have gone over different scenarios regarding which members are in compression and which are in tension; the only time we did that was with Dr. Frame
- it was taught very well, the lab time was pointless
- the explanation I received on trusses from both teachers was very infomative, but when I used to software I didn't really know what I was doing or how to use it so it didn't help me understand trusses better…. After doing some of this test it would have been very useful to do an overview of material because I have forgotten alot of the info I learned about trusses. if i had been told to look over my old notes for 20 min before taking this test i would have done a lot better I believe. i'm assuming from the instructions that i should NOT look over any notes or materials before taking this test, so i hope this assumption was correct.
- it was a good software but it would have been nice to use it more than once
- it was a useful tool
- I think the software helped once you learned the software. So I think you should have 2 class periods instead of just one to get the full benefit of it, b/c it took 1/2 of the class to get the hang of it.
- it was nice
- my favorite part of the class was trusses b/c I understand it best and got a 100
- I had no problem w/ trusses.
- trusses was one of the easier sections in the class
- I think trusses were my favorite topic in the class. It think the information was taught very well.

- I think the software would have been helpful if you had more time to use it. The one class we had with it did not help in finding whether these forces were in comp, tension, because we were given only one day to use it.
- Trusses are fun.
- When I entered the test on trusses, I felt as though I had a good grasp on the material but it turned out to me lowest test grade thus far.
- I would spend a great deal of time studying, but on the tests I would have difficulty conveying my knowledge.
- Lots of work in this class
- The book never really explained a lot of steps in solving problems. Not trying HW probs before or after class led to a lot of confusion before studying on tests.
- I did the worst in this section and I think it may be due to the substitute teacher.
- Trusses were extremely fun to me once I got the hang of them. Thank you for sharing your knowledge with me.
- The material about trusses was taught in an easy manner which I liked and appreciated.
- Trusses are awesome.
- I found it more helpful to work through a problem on the board then do another one on my own.
- The tests should have more problems that are less in depth to help students display their in depth understandings better.
- The section on trusses was well taught, although at times I felt that the instructor's pace was a bit too fast when explaining the theoretical portions.

REFERENCES

Aminmansour, A. (1996). Utilizing information technologies to better educate engineers of tomorrow. In J. Vanegas & P. Chinowsky (Eds.), *Proceedings of the Third Congress on Computing in Civil Engineering*, pp. 362-368, New York: ASCE.

Angelo, T.A. & Cross, K.P. (1993). Classroom Assessment Techniques, Jossey-Bass, San Francisco.

Angelides, D. C., Poulopoulos, A., Avgeris, I., & Haralampous, P. (2000). Case studies and information technology in civil engineering learning. *Journal of Professional Issues in Engineering Education and Practice*, 126, 125-132.

ASEE (2002). ASEE profiles of engineering and engineering technology colleges, 2002 edition, available WWW: http://www.asee.org/colleges/.

ATIR Engineering Software. (2002). STRAP, software.

Ausubel, D. P. (1968). Educational Psychology: A Cognitive View, Holt Rinehart, & Winston, New York.

Ausubel, D.P. (2000). The Acquisition and Retention of knowledge, Kluwer Academic Publishers, Norwell MA.

Baker, N., Brent, R., Felder, R., Holzer, S., Miller, T., & Vouk, M. (1999). Effective teaching with technology, workshop presented at Clemson University, February 19.

Bedford, A. & Fowler, W. (2002). Working Model Simulation CD, software, Statics Study Pack, Prentice-Hall, New Jersey.

Bedford, A. & Fowler, W. (2001). Companion Website, WWW: http://www.prenhall.com/bedford.

Beer, F.P. & Johnston, E.R. (2000). New Media Version, software, McGraw-Hill, Columbus OH.

Bertz, M. (1998). Situated Learning Methodologies and Assessment in Civil Engineering Structures Education. Unpublished Dissertation, Georgia Institute of Technology, School of Civil and Environmental Engineering.

Biggs, J.B. (1999). Teaching for Quality Learning at University, St. Edmundsbury Press, Bury St. Edmunds, UK.

Bouchlaghem, N., Sher, W., & Beacham, N. (2000). Computer imagery and visualization in civil engineering education. *Journal of Computing in Civil engineering*, 14, 134-140.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning, *Educational Researcher*, 18, pp. 32-42.

CADRE Analytic. (2002). CADRE Lite, software.

Comer, P., & Geissler, C. (1998). A methodology for software evaluation, *Proceedings of SITE 98*: Society for Information Technology & Teacher Education, Washington, DC.

Computers and Structures, Inc. (2002a). ETABS, software.

Computers and Structures, Inc. (2002b). SAP 2000, software.

CSC (UK) Ltd. (2002). P-Frame, software.

Daily, B., Benjamin, C., & Riordan, C. (1992). Advanced technology classroom for engineering education. *International Journal of Applied Engineering Education*, 8, 56-60.

Daily, B. (1994). Multimedia and its impact on training engineers. *International Journal of Human Computer Interaction*, 6(2), 191-204.

DeVore, M. (2000). Statics Tutor, software, Prentice-Hall, New Jersey.

Dick, W. (1992). "An instructional designer's view of constructivism, Instructivism and the Technology of Instruction: A Conversation, Duffy & Jonassen editors, Lawrence Erlbaum Associates, New Jersey.

Dr. Software LLC. (2001). Dr. Frame, software, available WWW: http://www.drframe.com.

Durham, A.G., & Emurian, H.H. (1998). Learning and retention with a menu and a command line interface, *Computers in Human Behavior, 14,* 597-620.

Dymond, R. (1996). WWW and multimedia in undergraduate civil engineering. In J. Vanegas & P. Chinowsky (Eds.), *Proceedings of the Third Congress on Computing in Civil Engineering*, pp. 362-368, New York: ASCE.

Elshorbagy, A., & Schonwetter, D. J. (2002). Engineer morphing: Bridging the gap between classroom teaching and the engineering profession. *International Journal of Engineering Education*, 18(3), 295-300.

Felder, R.M. & Brent, R. (2003). Learning by doing. *Chemical Engineering Education*, 37(4), 282-283.

Felder, R.M., Felder, G. N., & Dietz, E. J. (1997). A longitudinal study of alternative approaches to engineering education: Survey of assessment results. *Proceedings of the Conference on Frontiers in Education, 1997*, 1284-1289.

Gattis, K.W. (2000). Long-term knowledge gains due to supplemental instruction in college chemistry courses, *Journal of Research and Development in Education*, 33(2), 118-126.

Georgia Tech – CASE Center (2002). GT STRUDL, software.

Glennan T. K., & Melmed A. (2000). Challenges of creating a nation of technology-enabled schools. The Josey-Bass Reader on Technology and Learning, pp. 48-79, San Francisco, Josey-Bass Inc.

Gramoll, K. (2001). Engineering Media Lab, WWW: http://ecourses.ou.edu/

Gramoll, K. (2002) Multimedia Engineering Statics, WWW: http://www.ecourses.ou.edu/statics/content_st/home.htm.

Guarino, J. (2001). Working Model Simulation CD, software, Statics Study Pack by R.C. Hibbeler, Prentice-Hall, New Jersey.

Hays, W.L. (1994). Statistics, Harcourt Brace, Fort Worth, TX.

Healy, A.F., Clawson, D.M., McNamara, D.S., Marmie, W.R., Schneider, V.I., Rickard, T.C., Crutcher, R.J., King, C.L., Ericsson, K.A. & Bourne, L.E. (1993). The long-term retention of knowledge and skills. *The Psychology of Learning and Motivation*, 30, 135-164.

Hein, M. F., & Miller, S. (1995). Modern spreadsheets: Dynamic learning tools for structures. In J. P. Mohsen (Ed.), *Proceedings of the Second Congress on Computing in Civil Engineering*, pp. 943-946), New York: ASCE.

Heller, R. S. (1991). Evaluating software: A review of the options. *Computers & Education*, 17(4), p. 285 – 291.

Hesketh, R.P., Farrell, S. & Slater, C.S. (2003). An inductive approach to teaching courses in engineering. *Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition*, session 2351.

Hibbeler, R.C. (2001). Companion Website, WWW: http://cwx.prenhall.com/hibbeler.

Hibbeler, R.C. (2002). Engineering Mechanics. Statics, Prentice-Hall, New Jersey.

Hmelo, C.E., Lunken, E.Y., Gramoll, K., & Yusuf, I. (1995). Multimedia courseware for teaching dynamic concepts: Assessment of student learning. *Proceedings of the 1995 Frontiers in Education Conference*, 2b2.19-2b2.23.

Holzer, S. & Andruet, R. (2000) Statics Learning Environment, software, available WWW: http://www.succeednow.org/content/Statics/index.html.

Iano, J. (1998). Shaping Structures: Statics, software, John Wiley & Sons, New York.

IES, Inc. (2001). Visual Analysis, software.

ISDC, (2002). BEST Statics, WWW: http://web.umr.edu/~bestmech/preview.html.

Issa, R. R. A., & Domitrovic, R. C. (1995). Multimedia-based instruction for engineering mechanics courses. In J. P. Mohsen (Ed.), *Proceedings of the Second Congress on Computing in Civil Engineering*, pp. 943-946, New York: ASCE.

Issa, R. R. A., Cox, R. F., & Killingsworth, C. F. (1999). Impact of multimedia-based instruction on learning and retention. *Journal of Computing in Civil Engineering*, 13, 281-290.

Jones, A., Scanlon, E., Tosunoglu, C., Morris, E., Ross, S., Butcher, P., & Greenberg, J. (1999). Contexts for evaluating educational software. *Interacting with Computers*, 11, P. 499 – 516.

Kenimer, A. & Morgan, J. (2003). Active learning exercises requiring higher-order thinking skills. *Proceedings of the 2003 American Society for Engineering Education Annual Conference and Exposition*, session 1237.

Kolodner, J.L. (1997). Educational implications of analogy: A view from case-based reasoning", *American Psychologist*, 52(1), pp. 1-10.

Krone, S. (1995). Instructional aids: Preferences and influences on achievement. In J. P. Mohsen (Ed.), *Proceedings of the Second Congress on Computing in Civil Engineering*, pp. 943-946, New York: ASCE.

Kulik, C.C. & Kulik, J.A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior*, 7, 75-94.

Lalley, J.P. (1997). Comparison of text and video as forms of feedback during computer-assisted learning, *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 58(1-A), 84.

Lalley, J.P. (1998). Comparison of text and video as forms of feedback during computer assisted learning, *Journal of Educational Computing Research*, 18(4), 323-338.

Lenox, T.A., O'Neill, R.J., & Dennis, N.D. (1995). Computer integration throughout an undergraduate civil engineering curriculum. In J. P. Mohsen (Ed.), *Proceedings of the Second Congress on Computing in Civil Engineering*, pp. 943-946, New York: ASCE.

Mayer, R.E. & Chandler, P. (1993). "When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology*, 93(2), pp. 390-397.

Mayer, R.E. (2001). Multimedia Learning. Cambridge University Press, Cambridge, UK.

McDougall, A. & Squires, D. (1995a). A critical examination of the checklist approach in software selection. *Journal of Educational Computing Research*, 12(3), p. 263 – 274.

McDougall, A. & Squires, D. (1995b). An empirical study of a new paradigm for choosing educational software. *Computers & Education*, 25(3), p. 93 – 103.

McGill, D. J. & King, W.W. (1995). Engineering Mechanics, Statics, PWS, Boston.

McKeachie, W. J. (1999). McKeachie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers, Houghton Mifflin, Boston.

Meriam, J.L. & Kraige, L.G. (2002). Engineering Mechanics Statics, John Wiley& Sons, New York.

Meyer, K. F., & Ressler, S. J. (1995). A tool but not a crutch: Using design software in an undergraduate structural steel course. In J. P. Mohsen (Ed.), *Proceedings of the Second Congress on Computing in Civil Engineering*, pp. 943-946, New York: ASCE.

Miller, G. R. & Cooper, S. C. (1995). Something old, something new: Integrating engineering practices into the teaching of engineering mechanics. *Journal of Engineering Education*, 84(2), 105-115.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 12(2), 177-213.

Moshell, J. M. & Hughes, C. E. (1996). The virtual academy: A simulated environment for constructionist learning. *International Journal of Human Computer Interaction*, 8(1), 95-110.

MSC Software. (2003). <http://www.workingmodel.com>.

Muscarella, J. V., O'Neill, R. J., & Gano, J. A. (1997). STACKER: A visual basic tool to help teach the direct stiffness method. In T. M. Adams (Ed.), *Proceedings of the Fourth Congress on Computing in Civil Engineering*, pp. 281-288, New York: ASCE.

Neisser, U. & Hyman, I. E. (1999). Memory Observed: Remembering in Natural Contexts, Worth Publishers, New York.

Office of Research Compliance (2001). Institutional Review Board, WWW: http://www.osp.gatech.edu/compliance/humans/about_irb.html

Orey, M., Zhao, R., Fan, H., & Keenan, R. (1998). Summative evaluation of the SINCGARS Tutor. *Computers in Human Behavior*, *14*(4), 579-595.

Palloff, R. M. & Pratt, K. (2001). *Lessons from the Cyberspace Classroom: The Realities of Online Learning*. San Francisco: Josey Bass.

Papert, S. (1991). Situating constructionism, Constructionism, Harel & Papert editors, Ablex, New Jersey.

Petraglia, J. (1998). Reality by Design: The Rhetoric and Technology of Authenticity in Education, Lawrence Erlbaum Associates, New Jersey.

Philpot, T.A. (2002). MDSolids, software, available WWW: http://www.mdsolids.com.

Premkumar, K., Hunter, W., Davison, J., & Jennett, P. (1998). Development and validation of an evaluation tool for multimedia resources in health education. *International Journal of Medical Informatics*, 50, p. 243 – 250.

Research Engineers International. (2002). STAAD Pro, software.

Reeves, T. C., & Harmon, S. W. (1994). Systematic evaluation procedures for interactive multimedia for education and training, Multimedia computing: Preparing for the 21st century, S. Reisman editor, pp. 472-505, Idea Group Publishing, Harrisburg, PA

Riley, D. R., & Pace, C. (1997). Improving teaching efficiency with multimedia teaching aids. In T. M. Adams (Ed.), *Proceedings of the Fourth Congress on Computing in Civil Engineering*, pp. 281-288, New York: ASCE.

RISA Technologies. (2002). RISA-2D, software.

Ruthkosky, K.O. & Dwyer, F.M. (1996). The effect of adding visualization and rehearsal strategies to advance organizers in facilitating long term retention. *International Journal of Instructional Media*, 23(1), 31-40.

Samek, J. V., & Landry, K. A. (1997). Dynamic visualization in the classroom using computer simulation. In T. M. Adams (Ed.), *Proceedings of the Fourth Congress on Computing in Civil Engineering*, pp. 281-288, New York: ASCE.

SAS IP, Inc. (2002). ANSYS, software.

Semb, G. B., & Ellis, J. A. (1994). Knowledge taught in school: What is remembered? *Review of Educational Research*, 64(2), 253-286.

Silverstein, D.L. & Baker, J.R. (2003). Improving retention of calculus by engineering students in small programs, *Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition*, session 2793.

Siegel, S. & Castellan, N. J. (1998). Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, New York.

Sirkin, R. M. (1999). Statistics for the Social Sciences, Sage Publications, Thousand Oaks, CA.

Son, B. & VanSickle, R.L. (2000). Problem-solving instruction and students' acquisition, retention, and structuring of economics knowledge. *Journal of Research and Development in Education*, 33(2), 95-105.

Spellman, G. (2000). Evaluation of CAL in higher education. *Journal of Computer Assisted Learning*, 16(1), 72-82.

Spiro, R.J., Feltovich, P.J., Jacobson, M.J., & Coulson, R.L. (1992) Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains, Instructivism and the Technology of Instruction: A Conversation, Duffy & Jonassen editors, Lawrence Erlbaum Associates, New Jersey.

SPSS Inc. (1998). SPSS for Windows, software.

Squires, D. & Preece, J. (1996). Usability and learning: Evaluating the potential of educational software. *Computers & Education*, 27(1), p. 15 – 22.

Squires, D. & Preece, J. (1999). Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. *Interacting with Computers*, 11 (5), 467-483.

Sulbaran, T.A. (2002). Impact of Distributed Virtual Reality on Engineering Knowledge Retention and Student Engagement, Unpublished Dissertation, Georgia Tech.

Sulbaran, T. A. & Baker, N. C. (2000). Enhancing engineering education through distributed virtual reality. *Proceedings of the Conference on Frontiers in Education, 2000*, p. S1D-13— S1D-18.

Thukral, M & Gramoll, K. (2002). Web portal for basic engineering courses with a problem database and integrated communications. *Proceedings of the 2002 American Society for Engineering Education Annual Conference and Exposition*, Session 2368.

Ubuz, B. (2001). First year engineering students' learning point of tangency, numerical calculation of gradients, and the approximate value of a function at a point through computers. *Journal of Computers in Mathematics & Science Teaching*, 20(1), 113-137.

Whitley, G.E. (1995). Principles of Research in Behavioral Science, Mayfield Publishing Company, Mountain View, CA.

Will, J.D. & Tougaw, P.D. (2003). An integrated pair programming experience. *Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition*, session 2420.

Williams, T.C. & Zahed, H. (1996). Computer-based training versus traditional lecture: Effect on learning and retention, *Journal of Business and Psychology*, 11(2), 297-310.

Yildirim, Z., Ozden, M.Y., & Aksu, M. (2001). Comparison of hypermedia learning and traditional instruction on knowledge acquisition and retention, *Journal of Educational Research*, 94(4), 207-214.

VITA

Sean St.Clair was born in Lewiston, Idaho and later raised in Nampa and Twin Falls Idaho. He always enjoyed math and science and excelled in those areas. This led him to pursue a career in engineering. While attending Ricks College, where he earned a Specialized Associates degree in engineering, he was mentored by some amazing professors who encouraged him to pursue graduate school. He transferred to Utah State University where he earned a Bachelors of Science degree in civil engineering with an emphasis on structures. While in Utah, he married his wife, Zauhna. His love of teaching, structural engineering, and computers led him to Georgia Tech where he met a professor who conducted research in the area of educational technology in engineering. Sean spent six years at Georgia Tech, where he earned both masters degree and a doctorate in civil engineering, studying both structural engineering and engineering education. While at Georgia Tech, he also had the opportunity of practicing as a consulting structural engineer with Starzer Brady Fagan Associates. Sean had the opportunity to teach while at Georgia Tech and sincerely enjoyed interacting with and educating bright young minds. More importantly, while in Georgia, Sean and Zauhna were blessed with two beautiful children, Tucker and Fiona, and are expecting another daughter in the near future. Sean plans on continuing his career in both academia and industry and maintaining a relationship with the engineering education community.