INVESTIGATING THE EFFECTS OF TEXT ANNOTATION ON STUDENT RETENTION RATES

Ron Brown, B.S.O.E., M.A.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

May 2003

APPROVED:

Roger Ditzenberger, Major Professor
John Todd, Minor Professor
Jeff Allen, Committee Member and Coordinator of the Program in Applied Technology, Training and Development
Jon Young, Chair, Department of Technology and Cognition
M. Jean Keller, Dean of the College of Education, University of North Texas
C. Neal Tate, Dean of the Robert B. Toulouse School of Graduate Studies

This Study examines the grades of students using study skill methods and those who do not. The experiment consists of giving the treatment group the opportunity to use well-known study techniques. The Control group could only read the material. Both groups were given ten minutes to read a pre-selected text. The text consisted of an 1,807 word lesson on the, “Technical Training Management System.” Each group was given five minutes to take a twenty item quiz. Fifty-five students in the control group were limited to only reading the material. Fifty-six students in the treatment group could choose between highlighting, note-taking, and underlining. The results of the test scores were compared using a t-test for dependent samples. One week later, the same students in each group were re-tested, using the same quiz they had taken earlier. Students had five minutes to review study material. Study material for the treatment group included the same material they had annotated earlier. The Results from each group was compared. Efforts were made to avoid potential flaws in previous studies, thereby producing more viable results.

Results of this study indicate there is no significant difference between the grades of students who use the aforementioned forms of text annotation and those who do not.
Copyright 2003

By

Ron Brown
ACKNOWLEDGMENTS

I would like to thank those who have helped me to achieve this goal. To my Lord and savior Jesus Christ, who hand guides me throughout life and especially this process. To my mother, who has shown me true courage and conviction. To Betty, Fernando and Tobi, my close friends who pushed me to keep working. To Dr. John Griffin, my traveling partner and friend whose help will always be appreciated. I must thank my fellow BIC instructors, especially Dean Prine, who helped me complete my experiment. Finally, I want to thank my gym buddies, Buk, Tom Whaylen and Abdul who never let me forget to work on my study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. REVIEW OF RELATED LITERATURE</td>
<td>17</td>
</tr>
<tr>
<td>3. RESEARCH METHODOLOGIES</td>
<td>51</td>
</tr>
<tr>
<td>4. FINDINGS</td>
<td>72</td>
</tr>
<tr>
<td>5. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS</td>
<td>78</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>83</td>
</tr>
<tr>
<td>REFERENCE LIST</td>
<td>93</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>III-A TTMS Questions</td>
<td>57</td>
</tr>
<tr>
<td>2.</td>
<td>III-B Kuder-Richardson Formula</td>
<td>61</td>
</tr>
<tr>
<td>3.</td>
<td>III-C Procedures for T - Test</td>
<td>71</td>
</tr>
<tr>
<td>4.</td>
<td>III-D With Text Annotation</td>
<td>74</td>
</tr>
<tr>
<td>5.</td>
<td>III-E Without Text Annotation</td>
<td>74</td>
</tr>
<tr>
<td>6.</td>
<td>III-F Scores Breakdown</td>
<td>76</td>
</tr>
<tr>
<td>7.</td>
<td>III-G Paired Samples Statistics</td>
<td>76</td>
</tr>
<tr>
<td>8.</td>
<td>III-H Paired Samples Correlations</td>
<td>77</td>
</tr>
<tr>
<td>9.</td>
<td>III-I Paired Samples Test</td>
<td>77</td>
</tr>
</tbody>
</table>
INTRODUCTION

Background

The U.S. Air Force has established a program to train selected personnel to become instructors in various career fields. The instructor candidates are considered subject matter experts in their particular career field. Military training is designed to be straightforward and easily comprehended. However, military training provides a lot of information in a short period of time. The training is designed to show these subject matter experts how to consistently and effectively deliver pertinent instruction to those new military members who are not subject matter experts. The training also provides insight into proper curriculum design and development and allows the instructor candidates to see how they impact the overall Air Force mission. By training new members, each instructor candidate insures that there are enough mission-ready technicians available to perform the jobs necessary to keep the Air Force functioning at an efficient rate.

The instructor training course introduces many new concepts and principles. Students are required to study a large volume of material and extract the key concepts and important information. It is the extraction of the pertinent information, or rather the best method to use to extract pertinent information, which forms the basis of this study.
Various training administrators have had differing opinions as to the best way to help students study for this course. Depending on who is in charge at the time, students may or may not be given “sufficient” study materials that would allow them to annotate or visually cue selected text that they deem as important. Some training administrators believe that students need training materials which they can use and keep indefinitely. This allows students the opportunity to annotate selected text, which should aid student retention of information and thus result in higher grades.

The opposing view is that student study material can be recycled. Inherent in this view is the reasoning that students do not need to annotate in the study materials because pertinent information will be pointed out during the normal process of instruction. According to this opinion, students may indeed annotate what they believe to be important information, but that does not mean it is important information. Ultimately, if the instructor delivers the instruction properly, the students have all they need and do not need to annotate. Grades will not suffer because the students can easily discern the pertinent content.

The instructor program suffers from a cyclic approach to how study materials are given to students. There is little empirical evidence, as it relates to military training and instruction, which supports either argument.

The potential wastes of study resources, or possibly inadequate distribution of study materials, suggest reasons why this question should be explored. If either of these scenarios exists, it is detrimental to the academic environment. The results of this study
may provide needed information concerning whether or not text annotation positively affects student grades.

Significance of the Study

Research reveals an abundance of studies about student learning skills and study techniques. There are several specific studies on the subject of whether or not text annotation is effective. The vast majority of the studies indicate that there is no significant learning gain for groups who use text annotation as a study method. If there are already numerous studies on this subject, why is there a need for this study? This study will eliminate the shortcomings of many past studies and consequently provide more objective and viable results. The problems found in past studies are discussed later in this study. However, a glimpse at past experiments shows problems with such matters as time constraints, normal classroom settings, and proper test instrumentation. Factors such as these could lead to the inaccurate reporting of results. This study will produce results, free from variables, which were not properly accounted for in past studies.

Countless hours are spent employing text annotation techniques to help isolate information and increase retention of selected academic material. It seems a given that students and teachers alike will promote the idea that such text annotation techniques are necessary for the optimal retention of academic material.
However, there is not a substantial amount of evidence to support the contentions that text annotation alone, is effective. “Despite the attention paid to teaching effective learning habits, very little evidence is available to suggest the relative effectiveness of the several techniques recommended” (Arnold, 1942, p.449). Could it be that the effectiveness of text annotation has been enhanced purely through word of mouth?

Ultimately, students must somehow obtain the information they need from the material they are studying. How they find that information and what they do with it once they find it are questions that are continually explored. This study specifically deals with the processing of information once located, differing it from studies dealing with how students learn. “The study of locating information in text is warranted given the prevalence of this task in the school system and the work place” (McGoldrick, Martin, Burgering, & Symons, 1992, p.1).

The technical training institutions within most branches of the military rely heavily on teaching students basic study skills such as text annotation and expend a lot of time and energy into suggesting its use. Most centers even go so far as to create literature and offer lessons on how to use these study skills. The thought behind providing this additional help is to allow students to learn skills, that will enable them to better retain information. However, the question remains as to the effectiveness of certain study skills in helping student retention. Are instructors wasting time teaching and promoting text annotation?

Research from this study should prove particularly important in providing more concrete answers about whether present practices should be continued. Cashen and Leicht
(1970) reported that some experiments in the past have been flawed and thereby have led to faulty conclusions: “With few exceptions, investigators have not systematically studied the isolation effect in settings which permit easy extrapolation of conclusions to formal educational settings” (p.484). This study has attempted to overcome the flaws in earlier studies and present findings that have more validity. The results of this study may cause technical training centers to rethink their emphasis on certain study skills techniques. Secondly, this study could alter the amount of time and money put into promoting certain study skills, such as text annotation. Results of this study could be used to insure that time and resources are spent wisely on study skills that help students better identify and retain information.

Theoretical Framework

The search for tools that will enable learners to acquire knowledge more effectively and at a more efficient rate has led many educators to introduce a myriad of technological wonders in the classroom. The advent of computers in the classroom has introduced a new world of interaction. Instead of simple rote memorization, learners can now benefit from a more visual style of learning. The introduction of electronic games and simulations are touted as promising accelerated learning inside the classroom. Even the U.S. military is using simulators to train pilots because of their effectiveness in terms of cost and usability. According to one researcher, “Advances in modeling and simulation
technology will support new or better ways to train U.S. military forces during the next century, but that will depend on future requirements and investments” (Goldiez 1999). Still, in its most basic sense, even the high tech of learning work to achieve one ultimate goal -- teach the learner to retain as much information as possible. If, through the use of these elaborate and expensive tools, educators and trainers are attempting to give students techniques that will enable them to better retain information, then the argument that the use of good study skills may be an extraordinarily easier and more cost effective means to achieve this goal may have some validity.

While few would argue that good study skills aid in knowledge retention, the question of the effectiveness of certain study skills is constantly debated. Text annotation is a study skill that draws frequent fire. Text Annotation, is a term assigned to any of a number of various methods used to identify a specified text in an effort to enhance learning. “Text refers to the textual features of learning materials which influence comprehension and memory” (Collins, 1994). The benefit of text annotation is simply stated in research done by Di Vesta and Gray (1972) “In a sense, the learner has taken the initiative necessary to put the material into long term store; through encoding, the learner has linked the material to his existing cognitive structure, he has made it meaningful” (p. 8).

This study limits the types of text annotation, to be explored to highlighting, note taking, and underlining. Although some students highlight text and take notes, underlining seems to be the predominant choice among learners. "In a survey of 299
college students performed by Annis and Annis, reading and underlining was more than twice as popular as any other study method, and more than six times as popular as reading only" (Marxen, 1996 p.88). A curious point found in the results of the Marxen study was his assertion that underlining may be popular because it functions as a relaxation technique or as a method of justifying the assigned reading material to manageable proportions. According to Nist (1993), “Of all the effective strategies available to college developmental reading students, annotating (Noting important ideas or examples in text margins) and underlining have the widest appeal among students” (1993). Blanchard (1985) added, "Common sense would dictate that underlining is popular because it is easy to use, it requires no training, it is used by everyone, and it reduces the size of the text to study and review" (p.197). There are some specific questions that have researchers at odds. In order to answer these questions, certain assumptions relating to the effectiveness of text annotation must be stated. In these assumptions and throughout this study, the terms Student and Learner are used synonymously when speaking of the person who receives the information. The discussion hinges upon the following assumptions:

1. Specific study skills relating to text annotation are absolutely essential to the learner's ability to master given academic material.

2. If learners are not given enough academic material to fully employ the techniques of text annotation, their academic performance will significantly decrease.
3. Learners will always request enough materials to allow them to employ text annotation or some form of text annotation.

4. The trainers’ tasks will be more difficult if students cannot utilize the techniques associated with text annotation.

5. Text annotation alone is an effective study skills approach.

6. Note taking, underlining, and highlighting are all effective in helping student retention.

Assumptions 1 through 4 come from experiments such as the one conducted at Purdue University by Rikards and August (1975). Idstein and Jenkins (1972) laid the groundwork for assumption 5. A number of studies, such as the experiment by Davis and Ansis (1979), tested the effectiveness of each study skill against each other. These assumptions, beliefs, and implications should be examined. Does it greatly diminish the students’ ability to learn if they are not allowed to highlight or underline text materials? What effect does simple note taking play in the students’ ability to learn? How much of actual learning is based on the learner’s ability to pull important information from the text versus the trainer’s ability to teach in such a fashion that the learner can identify key points? Finally, how does a learner know what information is important enough to annotate?

Based on recent research, many scholars promote the idea that learners should be given materials with the “appropriate” text annotated. Trainers express the belief that annotated text results in a saving of the time spent going over the material. This model
also suggests that learners appreciate not having to bother with extraneous information. Trainers may want to add to this idea by saying that isolating the important material makes it easier for the learner to learn. What many scholars may not have envisioned is that, by altering the text, the interest level of the student may have been inadvertently altered. Schraw (1994) stated that “previous research has manipulated interest by using different texts or altering information within a single text rather changing the way readers study a single text” (p.3). Could it be that there is been an accidental breakthrough in helping students to learn? Some would argue that, indeed, by design or by accident, text annotation has proven its merit. However, the overwhelming majority of research fails to espouse this particular line of thinking. According to Hershberger (1964), “Complex typography designed to highlight important lesson content does appear to reduce the amount of unimportant material the reader learns, but fails, nevertheless, to facilitate learning of the important, core content” (p.295). Much of the research supports these findings, but at least one study has argued that there is no clear-cut evidence one way or the other because past studies have inherent flaws. Cashen and Leicht’s 1970 study, which questioned whether isolated material was retained better than non-isolated material, was inconclusive because the variable of study time was not controlled. “Since there was no assurance that students in the underlined condition (students allowed to isolate text) did not spend more time on adjacent statements, than students in the nonisolation condition, the validity of the assumption could not be assessed” (p.485). The flaw in Cashens and Leicht’s study could be a problem for other studies. Another model
suggested that note taking had the effect that most trainers are looking for. The encoding of information associated with note taking forces the student to retain some information in both short-term and long-term memory. However, what information is being stored in memory becomes the prevailing question. If the information the learners receive is erroneous, then it does not matter whether they retain it or not, because it is incorrect. Consequently, students may indeed rely more on a good instructor than on good study skills. Potts (1993) reported, “Several recent investigations have suggested that students needed help with their notes, as even successful students may fail to record many ideas communicated by the lecturer”. A teacher can do such an effective job in delivering the instruction and guiding the learning that students know what information is important and can anticipate future trends in learning. The students are given such a good learning plan that they know where to focus.

This model, however, weakens the argument that pure text annotation of written material is effective. Some research even states that the activity of organizing information for storage into memory is really the reason some learners believe they have learned. According to one study, “Disorganized presentation of prose material may in some cases actually facilitate recall” (Arkes, Schumaker, & Gardner, 1976, p.544). The fact that a learner has to put the puzzle together to make some sense out of the information, forces retention of the information. This theory also weakens the idea that simply underlining or highlighting material are effective means of retaining information by suggesting that the more active the learners are in their studies, the better their retention.
Instructors who endorse the use of study skills tend to assume that students possess the needed skills. Erroneous assumptions about the students, ability to apply good study skills may indirectly and falsely lead to other erroneous assumptions about their effectiveness. Educators’ concerns about text annotation are not limited to whether or not this is an effective study method for students. The costs associated with printing enough material to allow students to use text annotation could become a major budget concern. Some students will take notes in the margins of their given materials. Is there a justifiable need to give students consumable materials for every class?

Statement of the Problem

Student use of study skills continues to be one of the main strategies used to assist learners in acquiring and retaining pertinent academic information. The use of text annotation as a study skill strategy is readily accepted in most academic areas. However, the acceptance and apparent belief that this strategy will yield positive and measurable results lack substantial and credible support. Resources such as time and money are put into exploiting a strategy that seems based more on opinion than fact.
Purpose of the Study

The purpose of this study was to determine whether the study skill strategy of text annotation actually produces a positive and measurable effect on student retention. Is there any significant difference in academic performance between learners who employ text annotation and those who do not?

Research Questions

Does the use of certain study skills strategies enhance learner retention of information and consequently result in better performance academically? Do learners who do not employ certain study skill strategies show decreased retention of information, resulting in poorer performance compared to learners who do employ certain study skill strategies? Are some study skill strategies more effective than others in aiding learner retention of academic information?

Hypotheses

Text annotation is hypothesized to increase student retention of
pertinent academic data. Therefore student grades will increase. The following non-directional research hypotheses are examined:

\[ H_{01} \]: There are no significant differences in the progress check scores between students who use some form of text annotation and students who do not use some form of text annotation.

\[ H_{02} \]: There are no significant differences, after 1 week, in the progress check scores between students who use some form of text annotation and students who do not use some form of text annotation.

**Limitations**

The United States Air Force does not consider race or ethnicity as a factor when recruiting or delivering instruction to Air Force members. The Air Force considers race only to ensure that discrimination does not exist (Griffith, 1998). The Air Force also uses stratified random sampling to ensure that the number of females and males selected for instructor duty is proportionate to the Air Force population as a whole. With these facts in mind, this study does not consider racial, ethnic, or gender in instructor candidates as a separate variable.
Delimitations

1. The sample of students was limited to the instructor candidates attending the 82 TRSS Training Squadron, Basic Instructor Course, Sheppard Air Force Base, Texas, from November 2001 to May 2002.

2. Senior enlisted personnel and commissioned officers were excluded from this study for two reasons. Higher ranking personnel usually attend the basic instructor’s course to familiarize themselves with the job that their subordinates must perform. Second, because the Air Force generally selects instructor candidates from the E – 4 to E – 7 rank demographic for instructor duty, the number of senior ranking personnel attending the Basic Instructor Course is low.

Definitions of Terms

Basic Instructor Course: A 5-week course of study which teaches basic computer skills, correct speaking techniques, presentation techniques, and curriculum development and implementation.

Block of instructions (or “block”): Module of instruction that makes up an Air Force technical course. For example, an Air Force course may be made up of nine blocks.
These blocks usually end with a block examination to ensure that students have a firm understanding of what was taught during the “block of instruction”.

**Control group:** The group of randomly selected students who did not use any text annotation techniques during the Basic Instructors Course.

**Course of study:** The entire 5-weeks Basic Instructor Course taught at Sheppard Air Force base, Wichita Falls, Texas.

**Enlisted or Enlisted members:** Air Force members who will serve as instructors or instructor supervisors. They serve in the airman and noncommissioned officer ranks.

**Highlighting:** Emphasize or make prominent.

**Metacognition:** Having knowledge (cognition) and having understanding, control over, and appropriate use of that knowledge (Collins, 1986).

**Note taking:** A brief written record. A written summary or outline of something that has been observed or studied.

**Objective:** A statement of the student’s learning goal identifying the level of learning and subject of the lesson; the description of the component of a domain of learning.

**Post test- only control group design:** A research design that uses random sampling to divide a sample into two groups. The treatment group was given an opportunity to annotate text, and the control group was not. Both groups were then evaluated after the completion of a specific objective.

**Progress check:** A formal measurement tool used to evaluate student knowledge
for one specific objective.

**Study skills:** Learned abilities for acquiring knowledge and competence. They are skills for learning and tools to be used, for the purpose of this study, as a solution for some of the learning problems students encounter, both in school and elsewhere in their lives.

**Text annotation:** A term assigned to various methods used to identify a specified text in an effort to enhance learning. Specifically, this study examines note-taking, highlighting, and underlining.

**Text:** The textual features of learning materials that influence comprehension and memory (Decker, 1994).

**Treatment group:** The group of randomly selected students who used text annotation techniques during the Basic Instructors Course.

**Typography:** The art or process of printing with type. The general character or appearance of printed material.

**Underline:** Mark with a line or lines underneath; underscore. To indicate the importance of; emphasize, as by stressing or italicizing
The effort to “highlight” important text for future learning falls into a realm of learning called metacognition. Metacognition has been defined as “having knowledge (cognition) and having understanding, control over, and appropriate use of that knowledge” (Collins, 1986). Research studies abound relating to how certain metacognitive strategies may or may not help learners to learn. One such metacognitive strategy involves the use of text annotation. Educators and trainers alike continue to ponder the question of the effectiveness of visual cues or “text annotation” on the learning process. Another term for text annotation was used in a research study by Hershberger (1964). Typographical cueing, according to Hershberger, describes “underlining, variation of size of type, and variation in color of ink used to identify the various content categories” (p.290). Does allowing students to utilize some form of text annotation to highlight key information really help a student’s retention?

Clearly, there are those who believe that text annotation is extremely effective and beneficial. Text annotation has many supporters among students and teachers alike. From a student perspective, students may feel that without some form of text annotation, they are likely to retain less and remember less over a period of time. This research only speaks specifically to whether or not text annotation actually helps. However, Hale
(1983) offered some insight about student concerns: “Students not only overestimated how much forgetting will occur within a 2 week period but they expect a disproportionate amount of forgetting to take place within the first day after learning” (p.713). It is not difficult to find people who extol what they believe to be “hard” evidence that text annotation is an absolute necessity in the learning environment. Many teachers point out comments from students stating that, without text annotation, success would have been unlikely. What may be missing, for whatever reason, are the comments which some students say which place no added value on text annotation. Conversely, it is safe to say that some people believe that text annotation is not effective at all. Underlining as a text annotation device is frequently the focus of many debates about text annotation. “At the center of most underlining research is the “Von Restorff” effect, a laboratory finding that the isolation of an item against a homogeneous background produces increased recall of that item” (McAndrew, 1983 p.103). One researcher stated, “Nonetheless, much of the research in this area has failed to show that underlining by the teacher or the learner has any significant effect on how students learn (Hartley, 1986 p.218). Researchers have typically struggled to explain why underlining is so popular. “Underlining groups have not consistently outperformed other study groups, or for that matter, groups that do not use study strategies at all” (Blanchard& Mikkelson, 1986 p.199). One point of interest, which emerges from the Blanchard and Mikkelson study is the assertion that once students underline the information, they can go back later and study only the isolated text, allowing the students more rehearsal or study opportunities. The reason this point is of
interest is that it seems to indirectly support the claim that the initial underlining does nothing to increase student retention. It is only when the student reviews the material later that retention occurs. The McAndrew study supported this point by stating that underlining is not as efficient in producing comprehension as note taking unless students take time to study the underlined material. Rickards & August (1975) pointed out that the act of text annotation itself is not effective without certain conditions. Some form of mental processing or assistance in finding the important facts is required to make text annotation effective. “Regarding intentional learning, Crouse and Idstein (1972) demonstrated that readers given underlined text segments recalled significantly more underlined (intentional) material than those not given any underlined segments” (p.860). According to Blanchard and Mikkelson(1987),“Regardless of the reasons, the key for most students to a successful search and selection process is the amount and depth of cognitive processing that takes place during underlining” ( p.197).

One paper critiqued previous studies and tried to point out ways to make future experiment results more valid. Some of the studies included Leicht and& Cashen, Fowler and Baker, Crouse and Idstein (1972), Rickards and August (1975) and Hershberger (1964). In their research, Hartley, Bartlett, and Branthwaite (1980) suggested several points that have contributed to invalid results being presented as fact. The researchers concluded that many of the studies, which have supported the use of underlining as a viable means to improve student retention were not conducted as rigorously as they could have been. Some researchers have apparently tested the effects of underlining on items
that have already been cued. In some studies it seemed unclear whether or not students were prevented from underlining even if they were in the control group. Some studies, such as that of Crouse and Idstein (1972), seemingly redesigned their experiment in order to show significant differences between those who do and those who do not underline. Specifically, the sample size was increased, the passage was increased in word volume, and the experimenters asked fewer questions. This study was critical of the way experimenters either did or did not control for time. Some experiments allowed students to work until they finished studying and testing. Other studies established a time and regardless of whether they were finished studying or testing, when time was up, the students had to stop. “If we reject from consideration those studies that tested only cued (or Adjacent) items and those that did not in some way equate the time taken, then we are left with no study that provides clear-cut findings in favor of underlining. (Hartley, Bartlett, & Branthwaite, 1980, p. 89). According to this research, long term retention following underlining has been overlooked. Only a few studies reported long-term retention of data. The time length to test students after using underlining ranged from 6 days to 5 weeks.

One question that continues to surface is whether or not one study skill strategy is better than another. While this research does not explore this question, some discussion of this topic seems appropriate. Which one of the three methods of study presented in this research is best? Reading is reported to be the most effective study skill strategy for both immediate and delayed recall, according to the discussion presented by Noall (1962):
“Reading was the only significant difference which existed among the three methods on immediate recall tests” (p.52).

In another study (Stordahl & Christensen, 1956), 200 Air Force basic trainees (airmen) were given two separate passages to read. The airmen were randomly assigned to one of four study technique groups: reading and underlining, reading and making an outline, reading and summarizing, and reading only. Time was not held constant. A pilot study was conducted using college freshmen. Based on the results of that pilot study, the mean times for studying the passage using the pre-selected study method was used as the allotted time for each group. For most groups it was a little less than 1 hour. The airmen were tested immediately and 1 week later. Although the subjects for this study were basic trainees who were similar in demographics, an analysis of covariance was used based on the Air Force pretest. The Air Force pretest or ASVAB test scores students in four basic categories: mechanical, electrical, general, and administrative. These scores are used to place students in a career field reflective of their aptitude. “The results were very consistent; none of the analyses showed significant differences among study techniques” (Stordahl & Christensen, 1956, p.569). The researchers noted that, without training in how to use the techniques investigated in this study, the techniques would be no more effective for a general Air Force population than simply reading and rereading.

In his research, Arnold (1942), studied the effectiveness of different study methods using 242 college students. The four techniques compared included: repetitive reading, underscoring and marginal note taking (consisting of writing on the text page), outlining,
and precis writing. Students were tested on excerpts from Latin American history. The study of the material was followed by the administration of an objective test. Five weeks later, the same test was administered again. According to Arnold (1942), “Rigid, identical time limits were maintained for all studying and testing in order to eliminate the factor of time variations” (p.452). The results of this study were generally consistent with results found in most other research. The researcher reported that his experiment showed no significant superiority of one technique over another. Reading and underscoring tended to be slightly more effective than other techniques, especially on immediate recall. The outlining technique showed a tendency toward the lowest scores.

In their research, Idstein and Jenkins (1972) tested 79 college students enrolled in an educational psychology course. Students could be randomly assigned to one of two groups: underlining or repetitive reading. The text material for both groups consisted of a 1,200 word passage containing information regarding procedures of a local government. Both were given 10 minutes of study time. After they finished, these groups were further broken down into groups involved in review sessions, which had 9 minutes or 4.5 minutes for underlining or groups which had 9 minutes or 4.5 minutes for repetitive reading. Students were immediately tested. Despite the fact that the groups with more review time had better grades, there was no significant difference found between those who underlined and those who simply reread the passage. The researchers went a little farther and tested another group of college students under the same basic premise, but this time, the subjects had to read a 6,000 word passage. Both groups had 50 minutes to
read the passage. Two weeks later the students were allowed 15 minutes of review and then were tested. The results were consistent with those found in the first experiment. Idstein and Jenkins concluded, that “apparently, underlining is no more effective than repetitive reading, even for long passages” (p.323).

Dyer, Riley, and Yerkovich (1979) conducted research, which examined separately, and in combination, notetaking, summarizing, and rereading when reading text material. The experiment included 191 juniors through graduate-level students attending Arizona State University. Three factors, Notetaking, Summarizing, and Rereading, were combined factorially to yield eight experimental groups. There was also one control group. Students were randomly assigned to one of the groups when they entered the experimental room. Once the students were seated they were given a envelope containing instructions, text, and supplementary materials appropriate to their condition. The experimental groups were given a 2,000- word passage describing a fictitious culture called the Himoots. The control group was given a 2,000 word passage relating to myocardial infarction. The posttest consisted of 30 completion questions. Each student received approximately 35 minutes to perform whatever function his or her group was assigned to do. According to Dyer, et.al, (1979), the major message of this study is threefold. College students will attempt to master the idea of a passage to a greater extent than factual learning. Different strategies, such as notetaking and rereading, are equally helpful in reaching the learning objective. However, and most importantly, any strategy that is used should be employed with the text material at hand for easy and constant
reference. Of all the topics introduced in the literature, none received greater attention than the encoding process (text annotation) itself. Specifically, much of the literature tries to point out what is actually happening during the encoding process (text annotation). The researchers have attempted to learn whether encoding processes are better than processes, which do not involve the encoding of information. Researchers have also tried to explain some of the misinterpretations of the results of some of the research about the encoding process.

Does it matter more if a student is given added instructions on how to properly use study skills methods, or is it more important that the student knows that the material presented will be tested later? This question again goes to the heart of the effectiveness of certain study skill strategies. Some research has indicated that there may not be enough training on the use of certain study skills methods to fully evaluate their effectiveness. In a study (Adams, 1969) done at Northern Illinois University, two classes enrolled in education courses were encouraged to underline material throughout the course. After comparing their grades to the grades of other classes of students, no significant difference was found. However, there was a concern raised in the study. While there was little correlation between underlining and improved grades, the research “provided some useful implications such as the need for more and varied research in the area and the need for systematic instruction in the underlining technique” (Adams, 1969). This “implication” would be tested.
In another study, (Craik & Martin, 1980), it was hypothesized that student grades would be better for students who were given a presentation on improving their underlining skills versus students who did not hear the presentation. The subjects were 77 students enrolled in an introductory psychology course. A single factor ANOVA was used for three groups: informed, uninformed, and control. The seminar leader lectured to the “informed about the test group” and to the “uninformed about the test group.” The control group listened to a lecture on *Psychological Abstracts*. Both the informed and uninformed groups would receive this lecture and associated material later in the course. Five weeks later, all three groups were tested with a multiple-choice exam relating to Psychological Abstracts.

Consistent with most of the literature, this study found no significant difference between students who listened to the presentation about improving their underlining skills and those who did not hear the presentation. The results of this study seemed to support the notion that just, because one has the tools, one doesn’t necessarily put them to good use.

“The results of the present study (Craik & Martin, 1980), indicate that just because subjects remember helpful suggestions does not necessarily mean that they can or will make practical application of the material” (p. 406). Probably the most significant comment by Craik and Martin study is their concern that while comprehension aids can be helpful, the results indicate there are other important influences on student behavior. This may be one of the reasons the study found a significant difference between the
groups who were told that the presentation material was testable and those who were not told the presentation material was testable.

In the Cashen and Leicht (1970) study, the researchers pointed out that “making materials conspicuous by underlining serves the same function as presenting the material as an advanced organizer” (p. 484). They hypothesized that the isolation of certain material would improve student retention and also the grades of their students. One point should be made about this study. The underlined material was selected by the experimenter and not by the student. Therefore, another variable was introduced into this study even though it seems it was unaccounted for. This study involved 40 freshmen enrolled in general psychology at Illinois State University. A control group and an experimental group were established. Students were randomly assigned to one of the two groups. For students in the experimental groups, five statements were underlined in the course-related text handed out to both groups. There was no apparent control for time in this research. Students were immediately tested using a multiple-choice test, which included one question on the isolated material.

The results of this research supported the idea that isolating an item may produce significantly improved student retention of material as measured by student grades. The retention of isolated material by the experimental group was found to be greater than the retention of that same material by students who did not have the information highlighted. However, it is significant that materials were underlined by the experimenter as opposed to being underlined by the student. In studies in which the student must find and
underline the important materials, rarely is there found a difference in grades. The researchers also noted that there was a possibility that the learning of the isolated material sometimes comes at the expense of the non-isolated material. However, researchers noted that since neither study time nor test time were measured nor held constant, “the validity of the assumption could not be assessed” (Cashen & Leicht, 1970, p. 485). In his study, Hershberger (1964) shed a bit more light on whether or not unimportant text would be lost while students spent more energy on “important” text. After studying 160 fifth-grade students, Hershberger reported mixed results. “Complex typography designed to highlight important lesson content does appear to reduce the amount of unimportant material the reader learns, but fails, nevertheless, to facilitate learning of the important core content (p. 295). One additional comment of author bears presenting. The researcher suggested that typographical cueing does not significantly affect reading times. Unfortunately, it does not appear that the author made a good attempt to account for time in this research; thus, his assumption warrants a great deal of scrutiny. Also, the maturity level of his population is a variable, which he seemingly made no attempt to account for as well. Despite quoting several studies from previous researchers who used mainly college-aged students, this researcher continued without any regard to the difference in the populations.

One concern of the studies, which is not evident in much of the literature is how students are tested. Hartley, et al (1980), pointed out that most students are tested using multiple choice, constructed choice questions, or both. While the researchers noted the
consistency of the testing method, they also pointed out that “it might well be that the failure to find significant differences results from the insensitivity of the measures used” (p.220). The main concern brought out by the literature is that frequently, the measurement device may be too limited to adequately evaluate the student’s knowledge. It appears that some researchers may have diminished the validity of their results by not asking enough questions. If students were asked too few questions, then there is a belief that they could either have used some form of deductive reasoning or simply guessed correctly to answer the questions. Quite simply, a measurement device with too few questions is not comprehensive enough to give adequate feedback about a student’s retention of certain information.

One last point to be made is that there is little knowledge on how children benefit or do not benefit from underlining. The majority of research has focused on college level students, and it could be that among children, underlining may make a more profound difference. As a result of their findings, especially relating to the lack of information on children, Hartley, et. Al, (1980) conducted an experiment using sixth graders as the participants. Fifty-five sixth graders, divided into two groups, took part in the experiment. One group was comprised of 30 “high ability” students, and one group was comprised of 25 “average ability” class. Students studied a passage containing 300 words. Two versions of the text were produced. The experimental version had 15 key words underlined, while the control version had none. The children were randomly assigned into the experimental and control conditions, based on alphabetical order of surnames.
Students were given a copy of the passage and a test sheet, which was placed face down. They were asked to read the passage through twice. When all the students had finished, they were asked to fill in the test sheet. Surprisingly, for both the reading and the test, there was no time limit. One week later, without warning, the students were retested.

The results of this study indicated that higher ability students did significantly better than children in the lower ability group. This result was consistent with studies of college level students. Children scored higher on cued words than on non-cued words. Scores on the cued words that were underlined in the experimental passage were significantly higher than scores on the same words that were not underlined in the control group. This last result indicates that experimenter-provided underlining may positively affect student retention. However, like studies involving college level students, do not support the claim that students who underline do better than students who do not.

Some researchers believe that many studies did not adequately account for the amount of time students were allowed to use when studying and using selected study skill strategies. Mathews (1980) set time limits of 25 minutes for a 2,000 word passage, which was read by 735 pupils. The total population of students was divided up into three distinct groups. One group could only read the material. The second group could read and underline. The third group was given the necessary materials to write down notes as they saw fit. According to the study, none of the differences in the means of each group was found to be statistically reliable. One of the more interesting findings was that the study method of reading only “consistently produced the largest scores both for factual
memory and ability to discover and reproduce the outline of the passage” (p.104). The researcher added, “It appears that when time was kept constant merely reading the materials with the intent to understand and remember was more effective than underlining and making marginal notes or outlining for immediate recall, and more effective than underlining and making marginal notes for delayed recall” (p. 106).

According to Marxen (1996), by not holding study times constant, researchers fail to note a seemingly obvious result. Different study skill strategies may require different amounts of time to accomplish. Therefore, students may find underlining a more effective study method because it forces them to spend more time on the material than if they simply read it without underlining. According to Marxen, “Several studies have shown that when subjects are given as much time as they need to study a passage in preparation for a test, they take longer to read and underline a passage than to read it alone”( p.88). In order to test his theory, Marxen conducted an experiment involving 160 college students. All students were supplied with a 908- word text reading from the Wall Street Journal. Students were divided into two groups. One group consisted of students who could only read the article, and the other group consisted of students who were allowed to read and underline the article. One point should be noted about the Marxen study. Students were randomly assigned to either of the groups without regard to their preferred study method. All students had to accomplish their reading of the article, using their assigned study skill strategy, within 12 minutes. Immediately after the 12-minute study period, all students were give a quiz. The results of the quiz scores indicated that, while holding time
constant, those that read and underlined did not score better than those who simply read the passage. Additionally, this research suggests that, “underlining may actually retard test scores compared to reading alone (p.91). In a survey to determine which of the 160 students used underlining, it was found that 104 students indicated that they “usually” or “always” underline. From this group, those who were not allowed to underline (50) outsored those who were allowed to underline (54). Regarding the last point, the researcher does note that since the sample size was so small, that findings may not be representative of general population. Under normal studying conditions, students are allowed to select the study method which best suits them. Some literature suggests that this may help students retain more information because they are more relaxed and comfortable using the study method. Since they do not have to learn another method, they can simply focus on learning the material.

Many of the same arguments for and against underlining can be used in the study of note taking. The reason note taking is considered in a study concerning text annotation is that it requires some of the same mental functions as text annotation. The learner must identify important facts and somehow encode this information to make retention easier. Note taking is more active and requires more effort than simple forms of text annotation. Does it make a difference if the learner takes notes or is given notes? Does encoding take place when one records the notes or when one goes back over your notes? Beecher (1988) summarized many research studies relating to encoding or cueing using note taking. There is no clear cut evidence to suggest that the act of note taking alone provides
any significant advantage for learning academic materials. It has been suggested that reviewing or organizing notes has added benefit, but these two functions are outside of simple note taking. Another study concluded, “There is a potential benefit to students from the encoding function of note taking when the lecture environment permits deep processing while taking notes that entail processing the information in the way they will be tested on it” (Anderson, 1983 p.657).

According to some educators, what really happens is that the instruction is good enough to lead them to the right direction of study or to the answer. So what some would call text annotation of important information is in fact, sound instruction. In fact, some believe that text annotation could be a detriment instead of helping. Marxen (1996) reported, “Unfortunately, by choosing to underline what they think is important, students may be essentially granting themselves a license to ignore the majority of the assigned materials” (p.94). The researcher also wrote that, while students achieve better scores on the underlined material, they underachieve on the material that has not been underlined. He concluded by saying that, based on the results of his study, “reading and voluntarily underlining parts of an unmarked text passage tends to decrease test scores on the passage, compared to the scores achieved by simply reading and rereading the passage” (p.93).

How can one effectively measure to see if true text annotation really works? This question may be hard to answer because of the ways research in this area has been conducted. According to Blanchard (1987), "Closer scrutiny revealed that much of this
confusion can be explained by the tangled and contorted web of methodologies used to investigate underlining” (p.197). If one highlights everything, does that mean one will remember everything? How do students know what material to highlight? “Several studies show that many students lack proficiency in identifying main points in texts, but little is known why students lack this skill (Schellings, Bernadette, Van Hout, & Gonny, 1995, p.742). The answer to the question of the effectiveness of text annotation has been and seems to be a source of continued debate among students and scholars. From a student perspective, they want everything they can possibly have to help them study more effectively and efficiently. If they have used the text annotation method, in any format, early in their educational development, they may now come to see it as an indispensable study tool. Some students may find simple note-taking and listening as the most effective study skill. Again, some mention should be made of any perceived difference between underlining and note taking. As stated earlier, for this study, text annotation will include underlining, highlighting, and note taking. Underlining and highlighting are so similar that any differences between the two seem very minor. Note taking?

Numerous studies attempt to provide data as they relate to the effectiveness of note taking. Two studies done by Di Vesta and Gray seem to be one of most referenced pieces of literature. In their first study, Di Vesta and Gray (1972) suggested that the only note taking that really aids students is the kind that serves a role in encoding. Some note taking serves external storage purposes. The latter note-taking purpose is usually
associated with only writing down the information and has limited-to-no processing. In the experiment, 120 subjects, which were divided into two groups, listened to three 5-minute passages. For each passage the overall procedure consisted of three segments: a 5-minute period in which the subjects listened to a recorded communication; a 5-minute interval; and a 3-minute testing period. The treatment group was allowed to take notes during the presentation. The control group was not. Both the treatment and control groups were further subdivided for the two treatments in the second interval. In the second interval, subjects were either allowed to rehearse the communication (by using their notes or by contemplating whatever they could remember of the message), or the subjects were prevented from rehearsing (by requiring them to work on a spatial relations test).

Both groups were tested immediately, using a free recall test, which scored for number of words and number of ideas, and a multiple-choice examination. The results are as follows:

1. The subjects who were permitted to take notes earned higher scores than those subjects who merely listened. The researchers admit that this result is open to speculation because the tests were so short.

2. When a 5-minute study period followed the listening period, a larger number of words were produced than when the study interval was filled by activities unrelated to the rehearsal of the passage. This fact once again adds weight to the influence of the encoding process versus just the external storage process.
3. The number of ideas recalled was favorably influenced by note taking, rehearsal, and testing.

According to Di Vesta and Gray (1972), “There were more words generated and higher multiple-choice test scores when the study interval was used for review than when it was used for other activities” (p. 8). From the research, it is clear that the authors believe that listening and note taking is more effective, however, they did not show clear evidence to support their theory. According to the researchers, “The number of ideas recalled was favorably influenced by note taking, rehearsal, and testing” (p. 8). What may be more telling is the comments from the researchers next study. In their second study, Di Vesta & Gray (1973), once again supported the importance of note taking as an aid to recalling material that one listens to: “Nevertheless, the reason for the beneficial effects of note taking is not answered by any of the studies, either the present one or any of those listed in the references (p. 285). Their second study acknowledged a point made in an earlier study, which stated that taking notes does not guarantee it being recalled at the time of the quiz. Yet another study suggested that, “one reason for the failure of the process of encoding notes to produce significant effects on later recall is that little encoding actually occurs” (McAndrew, 1983, p. 285). Armbruster and Anderson (1983) observed, “There is a potential benefit to students from the encoding function of note taking when the lecture environment permits deep processing while taking notes” (p. 657). The research
also noted that, despite the claims of some, note taking while listening does not interfere with learning.

Other studies have sought to explore and expound upon the experiments conducted by Di Vest and Gray. One such study (Fisher & Harris, 1974) included the question of note takings effect on immediate and later recall. In similar fashion to the Di Vesta and Gray studies, 112 subjects were selected and divided into a treatment and a control group. Each group was then further subdivided into groups combining different note taking and review combinations. A 40-minute lecture was followed by a 10-minute review period and a 30-minute testing period. A 15-minute posttest was given 3 weeks later. Consistent with most studies, “the results showed that a combination of taking notes and reviewing one’s own notes produced the most recall, while not taking notes and reviewing the lecture “mentally” produced the least recall” (Fisher & Harris, 1971, p.321). This result was supported by findings in a report on “Improving the Quality of Student Notes” (Potts, 1993). Recent investigations have revealed that even the most successful students may fail to record many ideas communicated by the lecturer. Results also suggest that “there is value in having students participate in the note taking process in that they are more likely to remember what they have noted, even if full notes are provided by the instructor” (Potts, 1993). A noted concern of Potts is that by providing students with full notes of the lecture, the students may not see the need for the lecture itself. This concern might be blunted by the fact that students often request annotated study material from the instructor. McAndrews (1983) stated that students express a marked desire for handouts
and say that handouts help them grasp the structure of the lecture more easily and allow them more time to concentrate. However, if this last premise has any truth to it, then the students themselves are advocating the notion that it is the time they have to review the notes which is most important, maybe even more so than the notes themselves.

According to the Beecher (1988) report, researchers have long debated the effects of note taking on student grades. However, most of the recent studies now acknowledge that “successful college student’s engage in greater integrative processing during note taking, and that note taking itself enhances organizational processing of lecture information” (Beecher, 1988). Yet another study pointed out that if note taking does nothing more than focus attention to the material, it could be suggested that note taking and underlining would achieve the same basic effect. This assumption is based on the premise that either one or both of these methods really are better than reading in the first place. In Kulhavy, Dyer, and Silver’s (1975) study, 144 high school students were randomly assigned to 16 per factorial cells. Students were given instructions on whether to underline, take notes, or do nothing based on the packet they received as they entered their respective classrooms. The experimental test was a 845 word passage titled, “The Island Battle”. Study time was not held constant, as each student took as much time as he or she thought necessary to complete the respective tasks. The posttest consisted of 60 questions with no time limit.

The results of the Dyer et. al (1979) study supported the earlier contention by Beecher(1988) that note taking is a more active process and therefore may help students
retain the information better than a more passive process such as underlining.

Unfortunately, the research notes that the higher grades gained by note takers were correlated with an increase in the amount of study time.

To be fair, there are studies which suggest that, even with note taking, the real learning does not occur until the learners can take time to process what they have written. Carter and Van Matre spoke to this issue in their 1975 study: “The benefit of note taking appeared to be derived from having an opportunity to subsequently review notes, and not from the act of note taking itself” (p.900). The implication from this study forced a rethinking of all forms of text annotation. Could it be that the review of any form of annotated text or identification of text deemed important is the real reason for information retention. Using this assumption, one can hypothesize that as long as the trainer directs students towards the right ideas and establishes a clear path to what is expected, the students can easily do the rest. Once learners have the “map” of where they are expected to go and what to do, they can easily decipher what information is considered important. This model would suggest that the trainer has much more to do with student retention of material than the activity students use to identify important information.

Critique of Existing Literature

The literature researched for this study provided a wide range of assumptions and concepts. It is was clear that some of the researchers wanted to prove their point, while
others may have not gone far enough to justify their findings. In some cases, the research was flawed and significantly decreased the validity of the findings. Most of the studies lent themselves to the notion that those study skill methods, which seem to be used universally, do not contribute to better retention of information. Consequently, according to the literature, it is more myth than truth that student grades suffer when students are not allowed to use certain study skill strategies.

The control of time given for review and testing is a question that should have been explored. Subjects were routinely tested using all forms of study skills methods. Unfortunately, the amount of time allotted for each study skill strategy was not controlled in most studies. Subjects who were selected or who themselves selected to use a study strategy other than simple reading rarely had a time limit in which to perform their task. Those subjects who used reading as their only strategy usually finished studying well ahead of the study strategy groups. By not controlling for time, certain studies could not have accurately reported results. Studies that did not control for time included Stordahl and Christensen (1956), Craik and Martin (1980), and Crouse and Idstein (1972). Those students who took longer to employ study methods other than reading may have been aided by the additional study time. Some may argue that inherent in allowing subjects to use different study strategies is the concession that more time will be granted. In a normal academic setting, course curricula are established prior to classes beginning, not as the class goes along. Inevitably, trainers try to design the course to allow students to process information. However, trainers do not distinguish between different processing and
encoding methods. Underliners receive no more time than note takers, who receive no more time than readers. When students are in class listening to a lecture, the lecture period is established by the amount of course content, not by how long the trainer thinks the students need to annotate material. By not duplicating this normal academic process when experimenting with certain study strategies, the researcher leaves open the question of the validity of his results.

How long a time period is it between studying and the testing of the students? Unfortunately, much of the research has done little to study the effects of immediate versus long-term testing. Davis and Annis (1978) made an attempt to speak to this concern in their study. If the question is whether or not certain study skills strategies affect student retention of information, then it seems inevitable that testing should occur both immediately and after some extended period of time. Most of the studies seemed content to test studies inside of an hour of the study period, with no additional testing later on. This omission of delayed testing is a glaring omission in any research study. Without delayed testing, the researcher cannot be sure that the results are indicative of students being recently associated with the material or whether or not actual encoding of the information happened. Once again, results from any study that bases its assumptions on immediate testing alone runs a high risk of being flawed. The researcher simply cannot say that the results are solely the results of a particular treatment and not the result of the unintended intrusion of another variable.
In conjunction with the question of time and whether or not researchers properly duplicate normal studying conditions, one other question comes to mind. Does it matter that students are selected to use a certain study skill strategy that may not be the one they normally use? Many of the studies set up experiments with a determined effort to randomize the selection of students into either the control or experimental groups. While this is good from a quantitative statistical point, some literature suggests that this may cause skewed results.

Several studies have sought to provide insight into whether or not certain study skill strategies aid student retention. Unfortunately, there also seem to be several different conclusions concerning the effectiveness of certain strategies. Many studies have suggested that it is of vital importance that researchers take into account the student’s preferred study skill method when experimenting with the effects of study skill methods. The possibility that preference may impact the results of this particular research question must be addressed in any future studies. Two studies by Annis and Davis explored the question of preference. In their (1978) study, Annis and Davis stated, “A possible reason for these inconsistent results is that subjects have been arbitrarily assigned to study conditions without any regard for the subject’s preferred method of study” (p. 175). The study involved 262 college students enrolled in 13 sections of a sophomore human growth and development course. Using a questionnaire, the students’ preferred method of study was determined. The group was divided in half, with one half allowed to use their preferred method and the other half used a nonpreferred method of study. Each group
was also divided into those who were familiar with the subject and those who were not. There were three subsections in both groups. In each group, students were part of either reading, underlining, or note-taking.

In this study, the researchers found that underlining seemed to be most effective with a familiar topic, but when it is the non-preferred method of study. On the other hand, note taking is most effective when it is the preferred technique but the subject is unfamiliar to the student. According to Annis and Davis, “The results of this study suggests that a student’s preferred method of study is very important and should probably be considered or controlled in future studies” (p.94).

In their 1976 study, Annis and Davis continued the argument that there is a definite positive correlation between the study skill method a person chooses and the success of the study skill strategy on student retention. They also contended that results from previous studies have varied widely, in part because of disregard for the students’ preferred study methods. The Annis and Davis study included a total of 120 students in a wide variety of academic areas. Students were asked to describe their study skill strategy. According to the survey results, 14% students used reading only, 39% used underlining, and 18% used reading and note taking. The remaining 29% used various other methods. Students were assigned to either their preferred or non-preferred study skill strategy. Students reviewed a 1,525-word article, once again under the auspices of either the preferred or non-preferred method of study. Students were also divided into two other categories. The specific methods of study used in this research were reading alone,
reading and note-taking, and reading and underlining. Half of the students were allowed 10 minutes to either reread their unmarked copy of the article or review their notes or outlines. The other half of the students, were not allowed to review prior to the examination.

The results of the Annis and Davis study indicated that the subjects scored best when they used a non-preferred study method and reviewed prior to the examination. One possible answer for these results is that when students use the non-preferred method of study, they seem to concentrate on the material more. They may actually spend more time with the material. A concept espoused by Annis and Davis was that, “the real value of underlining and note taking may lie in the fact that the use of these techniques requires the learner to spend more time with the learning material” (p. 177).

This concept was explored by Wade and Trathen (1989). In their study, 160 college students were allowed to use their preferred study skill strategy to prepare for an examination on a particular subject. The researchers had previously selected those ideas which they considered important, from the text. The text was a chapter on tides from *The Sea Around Us*, which had been reduced to 15 double-spaced pages of text. Students were tested individually. They were given all the necessary tools to accomplish either reading, underlining, or highlighting. Standardized scores on vocabulary and analogy tests were used to group subjects into categories of higher, average, or lower ability. Overall, the higher ability students remembered significantly more than the lower ability students, but not necessarily because they were better at using their pre-selected study...
method. According to Wade and Trathen, “These findings suggest lower ability college students do not have a problem identifying important information; however, they may have a difficulty focusing adequate attention on that information, or processing it well enough to learn it”(p.45). In this research students were able to select their preferred method of study. Would it have made a difference if the lower ability students were given a non-preferred study method?

During this research, lower ability students were asked to describe the study methods they used to identify and remember important information. It seemed that having lower ability students describe their preferred method of study helped “make them more aware of their studying behavior”(Wade & Trathen, 1989, p. 41). As a result, the student’s sensitivity to importance and recall increased. However, any possible increase in student retention would seemingly come from more focus on the material and not as a result of their using their preferred method of study.

The results of Wade and Trathen’s (1989) study reinforced the concepts postulated in the earlier studies that because students are able to use their preferred method of study, this will not necessarily translate into better retention. The researchers stated that “higher ability students learned significantly more information than did lower ability students”(p.45). Higher ability students learned more, not because of the study skill but rather, “effective studying involves for more than simply underlining or note – taking”(p.40).

Although this question arose in the literature, it does not seem to be as valid a concern as some might suggest. In assessing the impact of self-selected study methods versus
pre-selected methods, the researchers may not have been educated about the effectiveness of one study method compared to another. If the overwhelming majority of the literature is correct, there is no significant difference between one study skill strategy compared to another. Once again it bears pointing out that under normal study conditions, students function under the same conditions, thereby eliminating the possibility of one study group gaining an edge over another.

Some studies also seemed to have a problem with insuring that all subjects did what they were supposed to do. Most of the studies did make some attempt to implement control measures to insure the accuracy of their results, but some seemed oblivious to any problems. Studies, which compared one college class against another over the course of a semester or even longer, had to make the assumption that the students did as they were told. How could they possibly guarantee that those students who were supposed to read only did and those who could use a particular study method used the one selected for them.

One important issue routinely ignored by the literature is the impact of trainer-selected annotated text compared to student selected annotated text. This is clearly a problem, which impacts the results of many of the studies in this literature review. Students and teachers alike hail the use of certain study methods as a means to obtain better grades. Yet the literature pointed out repeatedly that the annotation is directly linked to the student’s ability to select pertinent information. Students cannot arbitrarily decide what is important and what is not. They need the guidance of the trainer. If some
students are receiving study materials which are already annotated by the trainer, then they are, in effect, being given less material to study. They can focus their efforts only on the facts that matter. Those students who must ascertain the facts and annotate those facts which they think are important are clearly at a disadvantage. Unfortunately, many of the studies, such as Cashen and Leicht’s (1970) study, made no reference to this point. Of all the concerns about the studies done in the past, this one may be the most significant, yet it is underrated. The results of any study, which uses pre-annotated text must be called into question. The students’ ability to find the important information, how much attention they give to information they deem unimportant, and the amount of time they must put into studying the material are all concerns that are ignored if the researcher fails to acknowledge this concept. As stated earlier, it appears that the studies reviewed in this research failed to adequately account for this variable and therefore diminished the validity of their results.

Along the same lines as the pre-selected text is yet another concern arising from the literature. Some experimenters, apparently not content with their initial findings, went back and altered the test to achieve the results they thought they should get. In a few cases, the researchers seemed so intent on proving their point that they lost their objectivity. As mentioned earlier, some researchers increased sample sizes, word passages were increased, and fewer questions were asked (Di Vesta & Gray, 1973). Several studies used large word passages and questionable measurement tools. Some studies had such poorly designed measurement tools that the results of their research may
well be indicative of the students’ ability to comprehend major points of the reading, but, the measurement device used failed to truly test whether certain study skills were more effective than others. In the majority of the studies which reported a significant difference between those who use certain study skill strategies and those who do not, this was a major problem.

The maturity level of the subjects is another issue that must be addressed in reviewing the current literature. Most of the studies used college-level students to test their particular hypotheses. However, some studies used students in academic levels as early as the fifth grade. The concern is whether the results of the studies, which used students in the early grades, can and should be generalized to more mature populations. Studies, which used a population consisting of academic levels of high school and below include Hartley et al (1980), Hale (1983), and Hershberger (1964). This issue is not as important as some of the ones mentioned earlier, but it does require some attention. Researchers should be careful when using the results of any study that uses students in the early grades. The experience of using certain study skills is greatly diminished. The ability to decide what information is important and what to annotate is also less in the students from the early grades. Based on the literature, one could make the case that certain study skill strategies may be more effective with younger students because of their academic maturity level. Note taking may be beyond their capacity to use effectively, thereby making underlining a more sought-after and “more effective” method for this group.

Does this necessarily mean that this trend will be seen in more mature groups? Because
of questions such as this, researchers need to be careful where they receive their data and how applicable these data are to their research.

While there may have been some concerns about the maturity level of the subjects being tested, all the researchers did do well in ensuring that the population that they sampled was similar. Only one study actually used a covariate analysis to account for potential differences between groups. However, to the credit of the researchers most of the studies seemed to go to great lengths to maximize randomization of the subjects, and the population as a whole were selected on specific demographics.

There appeared to be a great deal of literature on the topic of the use and the effectiveness of certain study skills on student retention of information. The studies, done correctly, posed the right questions, although did only what was necessary to present untainted results. Most of the researchers failed to do an adequate job in duplicating normal study conditions and testing in such a fashion as to ensure valid results. To some degree, many of the researchers seemed to be in a rush to report some data; they simply chose to ignore the problems of other studies. The repetitiveness of some of the flaws in many studies is remarkable. The lack of appropriate time controls or the inadequate control of the research environment were quite noticeable. The identification of some potential problems in past studies should have been a guard against flawed studies being done later. Even though most of the studies reported no significant difference between study skill methods, those in the minority seemed determined to prove their point. This determination may have come at the expense of valid testing and reporting.
Rickards and August (1975) used 90 college students to compare the grades of students who received experimenter-provided underlined material compared to students who could decide for themselves what to underline. “The group which underlined sentences entirely of their own choosing performed significantly better than any other group, and only they outperformed the reading only control group in total recall” (p.860). This study displays some of the major faults already mentioned; particularly, the study used a passage of only 80 sentences, and the students were only tested immediately. No delayed testing of the students was conducted.

Contributions of the Literature to Current Practice

Many of the studies had flaws that caused some of the reported results to be called into question. Researchers should take note of major flaws in previous studies and void the same problems.

Clearly, the discussion about placing universal time constraints on all of the subjects has not gone unnoticed. Based on previous studies, the present researcher employed strict time limits on both the immediate and delayed tests. For this research, each student had no more than a 5-minute review period prior to testing. Initially, students heard the lecture. One group was allowed to annotate the text as they wished, while the second group was not allowed to annotate at all. When the delayed testing occurs, students in both groups were given a 5-minute review period to either reread their notes completely or review their annotated material.
The fact that some studies elected to use both immediate and delayed testing is a major concept introduced into this discussion. This research employed both the immediate and delayed testing technique to overcome any concerns that the results of the study were applicable to the hypotheses being tested. Too many previous studies may have produced results that were more indicative of subjects remembering main points immediately after reviewing notes than whether or not student retention was improved with certain study skills methods. The comparison of student scores of the immediate and delayed testing is crucial to providing accurate, valid data. Any research that fails to use this important step is critically flawed.

In conjunction with the delayed and immediate testing concepts is the question of which tests yield the most valid results. In speaking about measurement validity and how it affects the subsequent data, one researcher made a good point: “It might well be that the failure to find significant differences results from the insensitivity of the measures used” (Hartley et al., 1980, p. 220). Depending on the size of the passage, the test could have too few or too many easy questions. The researcher must ensure that the measurement tools sufficiently measure what they are supposed to measure. Recent studies appear to have taken this point to heart, because more studies are using measuring tools that may give a more accurate look at student retention.

One other point that adds to the validity of the results is whether or not subjects were tested under normal study conditions. This research was conducted under normal operating conditions.
CHAPTER 3
RESEARCH METHODOLOGIES

Introduction

Research clearly supports the usage of quantitative statistics, ‘variables measured on the interval or ratio scales” (Hinkle, Wiersma, & Jurs, 1994, p.14). The data retrieved were number grades in the context of discrete variables. The research studies also seemed to favor the experimental design. Included in this experimental design was the use of inferential statistics. This particular study and others like it seem ideal for the experimental posttest only control group design (Campbell & Stanley, 1963). At this stage, this study wanted to learn whether certain study skills enhance learner retention, which in turn should lead to increased academic performance. The researcher did not make an effort find out why learners choose or choose not do use certain study skills. In short, the study wanted to determine what happens when a treatment is applied, not why the treatment was applied. This research will add to a large volume of research already done which utilizes the experimental approach.

Studies such as Idelstein’s, “Underlining Versus Repetitive Reading” Blanchard’s “Underlining Performance Outcomes in Expository Text” and Di Vesta’ and Gray’s (1972) “Listening and Note Taking” represent the thinking of most studies, and they all utilized the experimental design. This study fits all the criteria for choosing a quantitative strategy. It is hoped that this research can assimilate information from past studies,
couple it with insight gained from this study, and present new information concerning this topic.

Ethical Standards

Informed consent was obtained from the University of North Texas Human Rights Subjects Review Board. Written approval was obtained from the Commander, 82 TRSS, Sheppard Air Force Base, Texas, where the study was conducted. This report was made available to the 82 TRSS and the United States Air Forces in order to further knowledge regarding the effects of certain study skills on student retention. The study was designed to be as unobtrusive as possible, with its only impact being the identification of proven methods of study.

Population

The population was the classification of enlisted personnel selected for Air Force instructor duty. While there are some commissioned officers selected for instructor duty, their numbers are so few that they would have little if any impact on this research. The enlisted member is selected for instructor candidates school based on three key areas: rank, duty performance ratings, and education. The enlisted selectee is required to have at least an associates degree. Generally, most newly selected instructor candidates have an
associates degree but rarely anything beyond that. Most selectees are E-4s (sergeants/senior airman) and E-6s (technical sergeants). There are occasions when enlisted members of rank higher than E-6s are selected, but again their numbers are extremely small.

All selectees must have attained high ratings from their respective supervisors and the endorsement of their commander prior to being accepted for instructor duty. Selectees come from a variety of different work areas. Each of the selectees is considered an expert in his or her respective field. The instructor training they will receive will show them how to develop, present, and revise appropriate military training material. When selected for instructor duty, the selectees incur a minimum 4-year commitment to instruct in their specific job area.

Sample

The random sample consisted of 12 classes. Each class consisted of 8 to 12 instructor candidates. Randomization is one of the chief tenets of inferential statistics and is a critical way of neutralizing the possible effects of nuisance variables, greatly increasing the probability that the sample will be representative of the population as a whole and minimizing bias (Hinkle et al., 1994; Kirk, 1995).

Each class lasts between 20 to 24 days and has four units of instruction. There are three comprehensive exams, and multiple progress checks come after every unit except the last. Progress checks test one particular objective, while comprehensive exams test a
sampling of all objectives for that block. Based on fiscal year 2001 projections, classes should have 8 to 12 students each. The total number of students projected for the upcoming fiscal year is 535. This equates to approximately 179 students every 4 months. This exceeds the minimum of 20% of the population for an adequate sample size, as advocated by Gay (1987) for small populations. By allowing for more than the needed sample size, officers and senior enlisted personnel could be eliminated from the study without reducing the sample to an ineffective size. The role of the principal investigator was as participant observer. The investigator had to serve as a team member due to large number of trainees and a limited number of available trainers. The research did not require access to sensitive information such as social security numbers, thus avoiding major ethical concerns about privacy.

General Methods/Data Collection

This experiment used the posttest-only control group design (Campbell & Stanley, 1963). This design uses a treatment group (students allowed to use text annotation) and a control group (students not allowed to use text annotation). The design of this experiment can be expressed as follows:

Treatment group  R X O1
Control group    R  O2
Students were randomly assigned to a each Basic Instructor Course (BIC) class. There is no preplanned design for class composition. BIC instructors simply divide up the students from the total number of students on the projected arrival list. There is a total of 20 classes which were used for this experiment. Starting with 1 and ending with 20, all classes were given a number. All even-numbered classes formed the control group. This group was not allowed to use any study method other than reading the material. All odd-numbered classes formed the treatment group. This group was allowed to use one of the three study skills methods; underlining, note taking, or highlighting. Student materials were taken up after each class had completed that particular block of instruction. This allowed the researcher to see which study method was used.

Data-collection techniques included participant observation and spreadsheets denoting each student’s grades. Instructors documented grades on the appropriate forms for each appropriate unit of instruction. The principal investigator collected the grades from each instructor. The dependent variable for this research was the students’ progress check grades. The resulting data were quantitative in nature. Through use of the grade sheets, student’s scores were compared.

Military training has one common and basic tenet, consistency. Instructors are given specific guidance on what can be taught and how it can be taught. While there will be some differences in personalities among trainers, the rigidity of military training standards ensures the same instruction for every student.
Students in each group were given a copy of objective 6a, “Technical Training Management System” (TTMS). This objective contains 1,807 words of technical information and definitions. All students were given 10 minutes to study the material. All students were then tested. Results were recorded on the appropriate form. Students were given 5 minutes to take the tests.

One week later, at the end of the block, the same students were given 10 minutes to review their notes, highlighted material, or just reread the material. They were tested again using the same progress check. Students were given 5 minutes to take the tests.

The testing instrument was developed specifically for this study by the researcher. It consists of 20 multiple choice test items. All tests items were developed using the aforementioned TTMS guide. No other reference was used.

The progress check normally used to test this objective in the classroom was not used because it consisted of only five questions. By increasing the number of test items, the testing instrument becomes more comprehensive and consequently will yield more valid results. While increasing the number of test items is a step in the right direction, this step alone does not address concerns about the reliability of the progress check and each individual test item. A copy of the testing instrument is provided below:

III-A TTMS Questions
TTMS QUESTIONS

1. TTMS stands for:
   a. Technical terminology and Maintenance schedule
   b. Technical Training and Management System
   c. Technical transfer of Mathematical skills
   d. Technical Training and Maintenance System

2. TTMS tracks student information to the:
   a. Course and Class Number level
   b. Course, class, and block level
   c. Class, block, and objective level
   d. Block, objective, and test level

3. TTMS allows the instructor to:
   a. Check future student assignments
   b. Review enlisted force member recruitment contracts
   c. Create schedule instances
   d. Check all students date of rank

4. Instructors have the capabilities to perform all of the following except:
   a. Create an instructor record
   b. Give students credit for SIA
   c. View counseling documentation
   d. Give students credit for a test

5. The software that automates TTMS is known as:
   a. Plato
   b. Plateau
   c. Play-dough
   d. Plethora

6. The Air Force Training Management System (AFTMS) class number is:
   a. A five digit number that comes from your class roster
   b. A Six digit number that comes from your class roster
   c. An alphanumeric code which identifies your class in TTMS
   d. A 15 digit code which identifies your class in TTMS
7. The term which identifies tasks that the student has to complete during course instruction is the:
   a. Default Status
   b. PDSCLASS
   c. Schedule Instance
   d. Component

8. The term which typically identifies your flight is:
   a. Domain
   b. Organization
   c. PDSCLASS
   d. Qualification

9. The term which identifies your course and is 15 digits long is the:
   a. Scheduled instance
   b. Qualification
   c. Group instance
   d. Default status

10. The term that refers to a TTMS generated number which identifies a block of instruction is the:
    a. Group Instance
    b. Qualification
    c. Scheduled Instance
    d. PDSCLASS

11. The component that you will always choose to find a group of students is the:
    a. PDSCLASS
    b. Default Status
    c. Domain
    d. Qualification

12. The term organization in TTMS refers to your:
    a. Flight
    b. Group
    c. Wing
    d. Squadron

13. The term used to identify what a student did in relation to chosen component is the:
    a. PDSCLASS
b. Default Status 

c. Component 

d. Qualification 

14. **The term which identifies a class in TTMS and consists of the course number, AFTMS number, shift and section is the:** 

   a. Qualification 
   b. Group Instance number 
   c. Schedule instance number 
   d. Qualification 

15. **Which section acts as an human interface between AFTMS and TTMS?** 

   a. Military Personnel flight 
   b. Student Assignments 
   c. Registrars Office 
   d. Unit Personnel office 

16. **Who normally creates tracks day to day activities of the students using TTMS?** 

   a. Instructor 
   b. Instructor Supervisor 
   c. Course Superintendent 
   d. Course Supervisor 

17. **Who maintains instructor records within TTMS?** 

   a. Instructor 
   b. Instructor Supervisor 
   c. Course Superintendent 
   d. Course Supervisor 

18. **In the coming future, Courseware development for automated Instructional Systems Development (ISD) will become available in TTMS with what application software?** 

   a. Designers Edge 
   b. Windows 2001 
   c. Office 2001 
   d. Microsoft Outlook 

19. **As the Instructor, you will do all of the following in TTMS except:** 

   a. Create a Schedule Instance
b. Enroll students into a block
c. Record block test failures
d. Assign students to a course

20. **The term which refers to the student social security number is:**
   a. The student ID
   b. The PDSCLASS
   c. The Component
   d. The Domain

In an effort to ensure that the testing instrument was reliable, the Kuder-Richardson formula 20 for estimating internal consistency was used. Primarily, the KR20 procedure measures inter-item consistency. This formula estimates what the average reliability would be if one did all possible split half administrations. In evaluating the results of KR20, the researcher hoped for a high reliability coefficient. This in itself is no guarantee that the assessment is well suited to the outcome, but it can give one an idea of whether or not the items are strongly or weakly related.

In addition to providing data about the overall reliability coefficient of the testing instrument, each test item was measured to see how much “weight” each item added or subtracted from the overall test reliability rating. Cronbach’s Alpha (Also known as the internal consistency reliability coefficient) was used to determine the reliability coefficient of each test item.
The same testing instrument given to students in the experiment was given to 10 of the faculty development instructors, prior to the start of this study. The results of this procedure are as follows:

Table III – B  Kuder-Richardson Formula

<table>
<thead>
<tr>
<th>ITEM</th>
<th>X SCORE</th>
<th>X = X – MEAN SCORE – MEAN</th>
<th>X²</th>
<th>CRONBACH’S ALPHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>1.4</td>
<td>1.96</td>
<td>.6479</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>- 2.6</td>
<td>6.76</td>
<td>.1431</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>1.4</td>
<td>1.96</td>
<td>.6479</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>- 4.6</td>
<td>21.16</td>
<td>.6018</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>1.4</td>
<td>1.96</td>
<td>.6479</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>-.6</td>
<td>.36</td>
<td>.9945</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>1.4</td>
<td>1.96</td>
<td>.6479</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>1.6</td>
<td>2.56</td>
<td>.9958</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>- 2.6</td>
<td>6.76</td>
<td>.1431</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>1.6</td>
<td>2.56</td>
<td>.9958</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>1.6</td>
<td>2.56</td>
<td>.9958</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>-.6</td>
<td>.36</td>
<td>.9945</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>-.6</td>
<td>.36</td>
<td>.9945</td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>-.6</td>
<td>.36</td>
<td>.9945</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>1.4</td>
<td>1.96</td>
<td>.6479</td>
</tr>
<tr>
<td>16</td>
<td>10</td>
<td>1.4</td>
<td>1.96</td>
<td>.6479</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>-.6</td>
<td>.36</td>
<td>.9945</td>
</tr>
<tr>
<td>18</td>
<td>10</td>
<td>1.4</td>
<td>1.96</td>
<td>.6479</td>
</tr>
<tr>
<td>19</td>
<td>9</td>
<td>1.6</td>
<td>2.56</td>
<td>.9958</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>1.4</td>
<td>1.96</td>
<td>.6479</td>
</tr>
</tbody>
</table>

Mean = 8.6  \quad \Sigma x^2 = 49.60

Variance \quad \sigma = \Sigma x^2/N - 1 = 49.60/(20 - 1) = 2.61

\rho_{KR20} = k/(K - 1) * (1 - M^2(k - M)/k\sigma^2)

\rho_{KR20} = 20/(20 - 1) * (1 - (8.6^2(20 - 8.6))/(20 * 2.61)

\rho_{KR20} = 1.05 * (1 - 8.6(11.4/52.20)

\rho_{KR20} = 1.05 * (1 - 98.04/448.92)
\[ \rho_{KR20} = .821 \ (\text{Reliability Coefficient}) \]

M – The Assessment Mean (8.6)
K – The number of Items in the Assessment (20)
\( \sigma \) – Variance (2.61)

Cronbach’s Alpha:
\[ \alpha = \frac{K}{(K - 1)} \times \frac{\sum s_i^2}{s_{\text{sum}}^2} \]

\( s_i^2 \) – Denotes the Variances for the K individual items
\( s_{\text{sum}}^2 \) – Denotes the Variance for the sum of all Items

One other concern regarding the testing instrument was that of construct validity. Were the methods used in this study sound enough operationally to warrant legitimate inferences from the results of this study? Particularly, was the testing instrument developed well enough to produce legitimate results that could be generalized to an entire population? The testing instrument in this study overcame any construct validity concerns.

When the construct validity concern is raised in regards to the measures used to test students, it is referred to as Mono-method Bias (Cook et al. 1979). The concern is that the researcher thinks the instrument is broad enough to get a comprehensive look at the item in question. In reality, the instrument only gives you a piece of the whole picture.

In this study, one way used to combat mono-method bias was to ensure comprehensiveness of the testing instrument. Only one objective was used as the item of
study. In past studies, the passages students had to read have sometimes became somewhat convoluted or they were one of many parts of study material.

One hard concept of military training is that students be trained and evaluated on one item at a time. At the end of a specific block of instruction, they may be given an accumulative test, but prior to that students understand that they can focus on only one thing before moving on. The methods used in this study duplicated the student’s normal daily activity. This included the testing procedure. Students are normally given a progress check at the completion of each objective. The only difference between the testing instrument used each day by the staff and the one for this experiment was that the testing instrument used for this study was longer. More questions were asked on the one objective than there normally would be.

Since a double blind procedure was used, the students did not know why there were more test items. They simply saw it as part of the normal process. This helped to reduce or eliminate possible evaluation apprehension.

The normalcy and consistency of testing within the military training environment breeds an expectancy for tests among the students. The increase in the number of test items on the testing instrument, the calculated reliability of the testing instrument, and no increase in the normal student material load help to ensure construct validity of the testing instrument.
Data Treatment Analysis

In deciding how to handle the statistical analysis of data, it was easier to use the generic formula used to establish parameters and the report results of many studies. The first step requires the identification of the null hypothesis or hypotheses. This study was attempting to determine whether or not certain study skill strategies affect student retention. Specifically, the null hypothesis states that there is no difference in grades between students who use certain study methods and those who do not. The independent variable was the study skills method (highlighting, underlining, or note taking). The dependent variable was student grades. Since there is only one independent variable, “the model is a simple regression” (Kirk, 1995, p.220). The goal of any educator is to help students to gain knowledge. This fact is as prevalent if not more prevalent in the realm of military training. U.S. military training must produce knowledgeable graduates in specific disciplines in a timely manner. While the military invests seemingly countless dollars in sophisticated technology, it must continually work to ensure that those personnel who must work on these systems can be trained quickly in an environment, which can be frequently chaotic. Many educators, both military and civilian, believe that the proper use of certain study methods can yield good results. If it is true that by using certain study methods student retention can be improved, then clearly these are questions the military or any educational organization needs to study.

The second step in the statistical analysis was to select the suitable statistical test or tests. Because it is hypothesized that there is no difference between the two correlated
groups, the variable is manipulated between the subjects and with the information known about the population, the test statistic was a between-subjects $t$ test for dependent samples. The tests draws on three statistics—the means, variances, and sample sizes to make statistical inferences about the significance of differences between sets of means. “The distribution of a statistic is called a sampling distribution” (Kirk, 1995) (p.51). The test statistic is based on the $t$-distribution, which tests the difference between means when the population standard deviation is not known. Unlike the sample statistic, which is used to describe characteristics of a particular sample or estimate population parameters, the test statistic specifically tests hypotheses about the population parameters. According to Kirk, the formula of the $t$ statistic is the mean of a sample minus the hypothesized value of the population mean, divided by the population standard deviation, which is divided by the square root of $n$ (the size of the sample used to compute the mean of the sample). In a more simplistic sense, $t$ equals the difference between 2 sample means divided by the standard error of the difference.

Subjects in the control and treatment groups were matched on three specific factors. All subjects were enlisted personnel, between the military rank of E – 4s (sergeant/senior airman) and E – 6s (technical sergeants). With rare exceptions, the subjects had 2 years or less of undergraduate education. All subjects are selected for instructor duty. Conversely, they must achieve and have maintained the exceptionally high personnel ratings. The population studied involved 12 classes, of no more than 12 and no fewer than 8 instructor candidates in the Basic Instructor Course.
In choosing an appropriate level of significance, several factors were considered. This particular study might well result in resource expenditure changes within the training section. However, the amount of money increased or decreased will not be substantial. This factor alone supports the use of a level of significance, which does not have to be extremely conservative such as one might find in medical research. “Traditionally, levels of significance set for rejection of the null hypothesis are at the .05 and .01 levels” (Thomas & Young, 1993, p.122). Kirk (1995) added that the most frequently used levels of significance are set at .05 and .01 levels. “Behavioral science researchers can often use higher alpha levels because of their concern about making type II errors” (Hinkle et al., 1994, p.170). Considering all these factors and the scope and importance of this research, the level of significance (alpha (α)) was set at, a priori, .05.

In selecting the level of significance, some concerns must be addressed. The researcher must be concerned with the power of the statistical test. According to Hinkle et al.(1994), power is the probability of rejecting the null hypothesis when it is false. In order to better understand power and what is needed in any good research, type I and type II errors must be explained. Any research must be concerned with rejecting a true null hypothesis (type I error) or retaining a false null hypothesis (type II error). Type II errors are commonly referred to as Beta (β). By minimizing β the power of the test is maximized (1 – β), but the researcher should not completely abandon any regard for making a type II error. Unlike α, β is not under the researcher’s direct control. However, the researcher still must do something to control β as much as possible. The researcher
could increase alpha, but that clearly leads to a less powerful test and an increased possibility of reporting and accepting invalid results. “However, suffice it to say that probably the best safeguard against the probability of a type II error is adequate sample size” (Thomas & Young, 1993, p.121). This study used an adequate sample size as a safeguard against a Type II error.

According to the literature (Kirk, 1995), several conventions are involved in hypothesis testing, such as the setting the probability of a type I error equal to or less than .05. Designing the experiment in such a fashion so that $\beta$ is equal to or less than .20 is yet another. By setting $\beta = .20$, the power of the test is $1 – \beta = .80$. Using the formula $p(\text{type II error})/p(\text{type I error})$ when $\beta = .20$ and the level of significance is set at .05, the aforementioned formula is in effect a ratio of $.20/.05$ (which equals 4). The researcher is saying that the type I error is four times more serious than a type II error. Another major point is brought out by Kirk. “A power of .80 is considered by many researchers to be minimum acceptable power” (p.59).

The hypothesis used in this research also contributes to the power of the statistical test. According to several researchers, including Hinkle et al. and Kirk(1994-1995) nondirectional hypothesis are more powerful than directional hypothesis.

Included in this step is the need to decide upon the distribution. “The distribution of a statistic is called a sampling distribution to distinguish it from the probability distribution” (Kirk, 1995, p.59). Since this study is using a $t –$ test, a $t$ distribution is used. The $t$ distribution is actually a family of distributions similar to a normal distribution.
They are symmetrical, bell shaped, and centered on the mean. The reason a \( t \) distribution is used instead of a normal distribution is because the \( t \) distribution changes as the sample size changes. “There is a specific (\( t \)) distribution for every sample of a given size” (Hinkle et al. 1994, p.183 ?). It is important to note that in order to select the proper \( t \) distribution for the experiment, it is necessary to understand the concept of degrees of freedom. The degrees of freedom is the number of observations minus the restrictions. According to Hinkle et al., this concept is important because, “each \( t \) distribution is associated with a unique number of degrees of freedom” (p.184).

In any experiment, the researcher must be concerned with factors, which may reduce valid inference making. They are “Undesirable sources of variation that affect the dependent variable” (Kirk, 1995, p.5). In short, one should reduce or eliminate those items, which become nuisances and bias the outcome of the study. Nuisance variables are a common factor in all experiments, but they can be controlled for. There are several options available for controlling nuisance variables. One means is to attempt to hold the nuisance variable constant for all subjects. In this research, all the subjects fit into a defined category. The pool of potential subjects included those enlisted members between the grades of E – 4 and E – 6. With rare exceptions, subjects had no more than 2 years of college. All subjects had between 10 and 15 years of active duty service time.

Randomly assigning subjects to the experimental conditions (Kirk, 1995) is one of the best methods of controlling nuisance variables. Randomization was a chief tenet of this research. By assigning subjects randomly, the researcher increased the chances of random
variation among observations. Groups were considered “equal” because their chance of being in either group was completely by chance. This particular option was used in conjunction with the first option. Two other specific measures used in this study to control nuisance variables were the Single Blind procedure and the disguised experiment technique described by Kirk (1995). A clear concern for this research was the demand characteristics, which could cause some students to alter their normal behavior in response to learning that they were a part of an experiment. Subjects were not informed about the nature of their treatment. More specifically, the subjects were not aware that they were participating in an experiment. One method for controlling nuisance variables is by using statistical control. “The analytic procedure employed in this statistical control is analysis of covariance” (Hinkle et al., 1994, p.483). According to Hinkle et al., the researcher measures variables in addition to the independent variables of primary interest and by controlling the variation attributed to these variables through statistical analysis rather than through research design. Because the subjects in this research were from such a defined population and because they were so evenly matched, statistical control was not used in this research. In most of the literature, researchers such as Blanchard and Mikkelson (1987), Marxen (1975), and Annis and Davis (1978) did not seem concerned about anything other than experimental control. Despite the fact that the test statistic ranged from t test to ANOVA, the researchers consistently avoided any talk of nuisance variables.
Treatment of the Data

The data retrieved from this research were straightforward. This is a basic tenet of quantitative data. The experimental research strategy contributes to careful control of the data by its very nature. Kirk (1975) stated the following: An experiment is characterized by the (1) manipulation by the researcher of one or more independent variables, (2) use of controls such as random assignment of subjects to experimental conditions to minimize the effects of nuisance variables, and (3) careful observation or measurement of one or more dependent variables. (p.6). The data were collected over a 6-month period. Once all the data were retrieved, an analysis of all the data took place. While some observations of the students were made, no reports or data resulted.

The results of the data were stored electronically in applicable software such as Word for windows.
Table III – Procedures for $t$-test

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) State the Statistical Hypothesis</td>
<td>$H_{01}: \mu_1 = \mu_2$ non-directional hypothesis</td>
</tr>
<tr>
<td></td>
<td>$H_{02}: \mu_1 = \mu_2$ non-directional hypothesis</td>
</tr>
<tr>
<td>2) Specify the test Statistic</td>
<td>$T$ Test for dependent samples</td>
</tr>
<tr>
<td>2) Specify the sample size and Sampling</td>
<td>Sample size of at least 107 (20% of 535)</td>
</tr>
<tr>
<td>Distribution</td>
<td>and a t distribution</td>
</tr>
<tr>
<td>4) Specify the level of significance</td>
<td>$\alpha$ is set .05</td>
</tr>
<tr>
<td>Obtain a Random sample of size (n),</td>
<td></td>
</tr>
<tr>
<td>Compute t, and make a decision.</td>
<td></td>
</tr>
</tbody>
</table>
The results of this experiment are discussed in this chapter. A brief review of the statistical procedure used, pertinent demographics about the population, and the hypotheses being tested are presented.

Initially the projected annual student population for the Basic Instructors Course was 744 students. As the fiscal year progressed, average class size consisted of only 9 students as opposed to the originally projected 12. The originally projected number of 744 students expected to be taught over the fiscal year was revised to 535 students. From this population, a sample of 111 students was drawn. A \( t \) test for dependent samples was used to determine differences between the two groups. The treatment group consisted of 56 students, who were allowed to select any of the three text annotation methods presented as an aid to help them study. The three methods allowed were highlighting, underlining, or note-taking. The control group consisted of 55 students, who only had the opportunity to read the information given to them. The scores from a predetermined testing instrument were analyzed between the two groups. A comparison of the two groups was made using the null hypothesis stated earlier.

Randomization was used to determine whether a student went into the control group or the treatment group. This was one of the ways used to ensure that the groups did not differ in any significant way. Students selected for instructor candidate school are
selected by the Air Force based on three key factors. These factors also help ensure that there is little variance among the demographics in this population. As stated, rank, duty performance ratings, and education levels are all factors in the selection process. The instructor candidates selected for instructor duty must be at least an E-4. An extremely small percentage of instructor candidates are over the rank of E-7, but this sector of the population generally serves in a supervisory position, not as full-fledged instructors. As such, the range of the rank selected is E – 4 to E – 7. Instructor candidates are required to have at least an associates degree or be within 1 year of completion. The average instructor candidate has approximately 2 years of college education. Instructor candidates must also be recommended for this position. Therefore, their duty performance ratings must be extremely high. In the selection process to become an instructor, even an average performance rating will disqualify a member from instructor candidate school. These three factors of rank, duty performance ratings, and education level coupled with the randomization used to place students in one of the two groups, helped to ensure that the students in the two main groups did not differ significantly from one another.

A breakdown of the groups is as follows:
Table III – D

With Text Annotation

<table>
<thead>
<tr>
<th>CLASSID</th>
<th>FIRSTSCORE</th>
<th>LASTSCORE</th>
<th># OF STUDENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Prine</td>
<td>13.34</td>
<td>16.00</td>
</tr>
<tr>
<td>2</td>
<td>Bryles2</td>
<td>14.00</td>
<td>17.00</td>
</tr>
<tr>
<td>3</td>
<td>List2</td>
<td>13.50</td>
<td>13.60</td>
</tr>
<tr>
<td>4</td>
<td>Reese1</td>
<td>12.00</td>
<td>16.00</td>
</tr>
<tr>
<td>5</td>
<td>Webber</td>
<td>13.80</td>
<td>16.30</td>
</tr>
<tr>
<td>6</td>
<td>Reese2</td>
<td>11.70</td>
<td>14.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>47</td>
<td>8</td>
</tr>
</tbody>
</table>

Total = 55 Students

Table III-E

Without Text Annotation

<table>
<thead>
<tr>
<th>CLASSID</th>
<th>FIRSTSCORE2</th>
<th>LASTSCORE2</th>
<th># OF STUDENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bryles1</td>
<td>13.33</td>
<td>14.77</td>
</tr>
<tr>
<td>2</td>
<td>Mac</td>
<td>14.25</td>
<td>16.63</td>
</tr>
<tr>
<td>3</td>
<td>List1</td>
<td>12.16</td>
<td>13.00</td>
</tr>
<tr>
<td>4</td>
<td>Leo1</td>
<td>13.13</td>
<td>13.63</td>
</tr>
<tr>
<td>5</td>
<td>Gunn</td>
<td>14.90</td>
<td>15.27</td>
</tr>
<tr>
<td>6</td>
<td>Leo2</td>
<td>10.50</td>
<td>12.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>43</td>
<td>13</td>
</tr>
</tbody>
</table>

Total = 56 Students

Hypothesis Testing H0₁: There are no significant differences in the progress check scores between students who use some form of text annotation and students who do not use some form of text annotation.
Hypothesis Testing \( H_{02} \): There are no significant differences, after 1 week, in the progress check scores between students who use some form of text annotation and students who do not use some form of text annotation.

In regards to \( H_{01} \), the critical value of \( t \) for a non-directional hypothesis, with 5 degrees of freedom and alpha set at .05 is +2.571. As shown in the results, the \( t \) value from this experiment was .290. Therefore, this study failed to reject the null hypothesis. Because alpha was set at .05, the probability of retaining a false null hypothesis is only 5%.

In more simplistic terms, the results of this study indicate that there is no difference between the test scores of students who use some form of text annotation and those students who simply read the material.

In regards to \( H_{02} \), the same groups of students were given the same testing instrument after a 1 week period. The same conditions applied for this experiment. The control group still did not use any form of text annotation, and the treatment group could not use any form of text annotation. The only difference is that students were tested after 1 week. The same testing instrument was used. All students spent time hearing a lecture on the subject they were initially tested on, but the lecture came on the same day following the initial test. This test was designed to see if there was any significance between the scores of the two groups after a sustained time period.

The critical value of \( t \) for this non-directional hypothesis, with 5 degrees of freedom and alpha set at .05 is +2.571. As shown in the table, the \( t \) value of this experiment was (-
4.321). This study failed to reject the null hypothesis. There was no significant difference between the scores of the two groups after a 1-week period.

The results of the experiment are outlined in Tables E through H:

Table III- F  Scores Breakdown

<table>
<thead>
<tr>
<th>Class With Text Annotation</th>
<th>Score for First Test</th>
<th>Score for Second Test</th>
<th>Class W/O Text Annotation</th>
<th>Score for First Test2</th>
<th>Score for Second Test2</th>
<th>Differ FsW – FsW/O</th>
<th>Differ D2</th>
<th>Differ Sqd D1</th>
<th>Differ Sqd D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prine</td>
<td>13.34</td>
<td>16.00</td>
<td>Bryles</td>
<td>13.33</td>
<td>14.77</td>
<td>.01</td>
<td>1.23</td>
<td>.0001</td>
<td>1.512</td>
</tr>
<tr>
<td>Bryles2</td>
<td>14.00</td>
<td>17.00</td>
<td>Mac</td>
<td>14.25</td>
<td>16.63</td>
<td>-.25</td>
<td>.37</td>
<td>.0625</td>
<td>.1369</td>
</tr>
<tr>
<td>List2</td>
<td>13.50</td>
<td>13.60</td>
<td>List</td>
<td>12.16</td>
<td>13.00</td>
<td>1.34</td>
<td>.60</td>
<td>1.795</td>
<td>.3600</td>
</tr>
<tr>
<td>Reese</td>
<td>12.00</td>
<td>16.00</td>
<td>Leo</td>
<td>13.13</td>
<td>13.63</td>
<td>-1.13</td>
<td>2.37</td>
<td>1.276</td>
<td>5.616</td>
</tr>
<tr>
<td>Webber</td>
<td>13.43</td>
<td>16.30</td>
<td>Gunn</td>
<td>13.91</td>
<td>14.91</td>
<td>-.48</td>
<td>1.39</td>
<td>.2304</td>
<td>1.932</td>
</tr>
<tr>
<td>Reese2</td>
<td>11.70</td>
<td>14.70</td>
<td>Leo2</td>
<td>10.50</td>
<td>12.23</td>
<td>1.20</td>
<td>2.47</td>
<td>1.440</td>
<td>6.100</td>
</tr>
<tr>
<td>SUM1</td>
<td>77.97</td>
<td>93.60</td>
<td>SUM2</td>
<td>77.28</td>
<td>85.17</td>
<td>-.31</td>
<td>8.43</td>
<td>4.804</td>
<td>15.6563</td>
</tr>
<tr>
<td>MEAN1</td>
<td>13.00</td>
<td>15.60</td>
<td>MEAN2</td>
<td>12.88</td>
<td>14.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table III-G

Paired Samples Statistics

<table>
<thead>
<tr>
<th>Pair</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Std. Error of the Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td>First Score</td>
<td>12.88</td>
<td>6</td>
<td>1.3701</td>
</tr>
<tr>
<td></td>
<td>First Score 2</td>
<td>12.99</td>
<td>6</td>
<td>.9209</td>
</tr>
<tr>
<td>Pair 2</td>
<td>Last Score</td>
<td>14.19</td>
<td>6</td>
<td>1.5735</td>
</tr>
<tr>
<td></td>
<td>Last Score2</td>
<td>15.51</td>
<td>6</td>
<td>1.3862</td>
</tr>
</tbody>
</table>
Table III-H

Paired Samples Correlations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>N</th>
<th>Correlation</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td>First Score &amp; First Score2</td>
<td>6</td>
<td>.705</td>
<td>.117</td>
</tr>
<tr>
<td>Pair 2</td>
<td>Last Score &amp; Last Score 2</td>
<td>6</td>
<td>.879</td>
<td>.021</td>
</tr>
</tbody>
</table>

Table III-I

Paired Samples Test

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error of the Mean</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td>-.1150</td>
<td>.9722</td>
<td>.3969</td>
<td>-1.1353</td>
<td>.9053</td>
</tr>
<tr>
<td>Pair 2</td>
<td>-1.3217</td>
<td>.7493</td>
<td>.3059</td>
<td>-2.1080</td>
<td>-.5354</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>Sig. (2 – tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td>-.290</td>
<td>5</td>
<td>.784</td>
</tr>
<tr>
<td>Pair 2</td>
<td>-4.321</td>
<td>5</td>
<td>.008</td>
</tr>
</tbody>
</table>
This study was done to evaluate the effects of text annotation on the grades of instructor candidates in the United States Air Force. This study came about because of questions raised concerning the funding of resources which would enable students to annotate given material, thereby eliminating the possibility of reuse by other students. Within the confines of the instructor candidate school, there seemed to be a consensus that supported the assertion that text annotation helped increase student grades. Based on a general, yet unsubstantiated theory, resources could be redirected to fund this idea. The basic premise of this study was that there is no difference between the grades of student who use text annotation and those who do not. This study could also point to areas for further studies relating to whether text annotation alone actually leads to better grades.

Previous research studies were reviewed, and the results of these studies helped ensure this study to avoid the flaws of past experimentation. Cashen and Leicht (1970) suggested that text annotation did in fact positively affect student grades. However, what may be more significant than their results is that this study showed how flaws lead to faulty assumptions. In the Cashen and Liecht study, there were no time constraints on study time. DiVesta and Gray (1973) went so far as to increase sample size; word passages were increased; and the number of test questions was decreased in order to show significant differences between the control and treatment groups.
Marxen (1996) seemed determined to eliminate potential flaws of past studies and report accurate data, no matter the outcome. His study showed no significant difference between the control and treatment groups, but more importantly, his study controlled for time, had ample sample size, and provided a sound testing instrument.

This study examined a random sample of 111 military instructor candidates. The entire population of instructor candidates consisted of 535 instructor candidates in one fiscal year. The instructor candidates were between E-4 to E-7 in rank, had approximately 2 years of college, and were all selected based on outstanding duty performance. Subjects were randomly assigned into control and treatment groups. There were 56 persons in each group. The strict Air Force criteria for selection as an instructor candidate and the random selection of students to either the control group or treatment group ensured that the groups were similar throughout the experiment.

The dependent variable used in this study was student grades. The independent variable used in this study was the text annotation method used by students in the treatment group.

Two null hypotheses were used to determine the effectiveness of text annotation on student grades. The experiment used the posttest-only control group design. Each hypothesis was tested using a $t$ test for dependent samples. A critical value of $t$ was established for both hypotheses.

Both hypotheses were retained. There was no significant difference between the test scores of students who use text annotation and students who do not use test annotation.
Secondly, after 1 week, there was no significant difference between the test scores of students who use text annotation and students who do not use test annotation.

Conclusions

Based on the findings of this study, the following conclusions were reached.

1. Findings of this study were consistent with earlier works by Stordahl and Christentsen (1956); Davis and Annis (1979; and Dale Marxen (1996). The overwhelming majority of the literature supports the conclusion that there is no significant difference between students who use some form of highlighting versus those students who do not.

2. There is no significant difference between students who use text annotation and students who do not, even over a sustained period of time.

3. The resources expended to allow students the opportunity to apply these specific study skill methods may be a poor investment. If there is no introduction of any other variables, giving student materials and telling them to highlight and study what they believe is important is a waste of resources.
Recommendations for Further Study

1. The results of this study were based on the premise that students did not receive any additional assistance in locating pertinent information to study. There could be a difference in grades between students who use text annotation and have been told what specific areas are important and those students who simply read the material, yet also know what specific areas are important.

2. The results of this study were confined to a select group of military personnel. However, the results of this study may be applicable to all levels of military members involved in military training. Clearly, studies need to be done on other predefined groups of military trainees. The results of this study suggest that, regardless of the rank and education level, as long as the treatment and control groups are not significantly different, the outcome will be the same.

3. Further research should be conducted to determine the best way to utilize text annotation. By itself, text annotation is not an effective study skill method, but coupled with other methods, it could prove effective.

4. Study skill methods which focus on understanding how to get the most out of reading material seem more applicable to the goal of increasing student retention. If text annotation is to be used, it will be most effective after the student learns how to extract pertinent content from a lesson. Even after the student is better able to glean
information from text, the text annotation method should be in conjunction with some type of lecture or outline for study.

5. Providing the students with an outline of the lesson content, complete with subcategories regarding which areas to study would help students focus in on important material. An outline would decrease the amount of time students would have to spend on less important information.

6. Instructor assistance and guidance will be extremely helpful in aiding student retention. Although, this method requires a great deal of instructor time, the Griffith (1998) study points out it’s clear benefit in aiding student retention.
Appendices
APPLICATION FOR APPROVAL OF INVESTIGATION INVOLVING
THE USE OF HUMAN SUBJECTS

University of North Texas Institutional Review Board for the Protection of Human
Subjects in Research (IRB)

1. Principal Investigator’s Name: Ron Brown

2. Home Address: XXXXXXXXXXXXXXXXXXX

3. Faculty Sponsor: Dr. Roger Ditzenberger   Phone: (XXX) XXX-XXXX
   This is in support of research for a Ph.D. dissertation


5. External Support is not required for this project

6. In making this application, I certify that I have read and understand the guidelines and
   procedures developed by the University for the protection of human subjects, and I fully
   intend to comply with the letter and spirit of the University’s Assurance and policy. I
   further acknowledge my responsibility to report any significant changes in the protocol,
   and to obtain written approval for these changes, in accordance with the procedures, prior
   to making these changes. I understand that I cannot initiate any contact with human
   subjects before I received approval and/or complied with all contingencies made in
   connection with that approval.

   Signature of Principle Investigator                                   Date

   Original Signed and Dated

7. Approval by Faculty Sponsor (Required for all Students): I affirm the accuracy of this
   application and I accept the responsibility for the conduct of this research and
   supervision of human subjects as required by law.

   Signature of Faculty Sponsor                                   Date
Original Signed and Dated

8. I have included copies of all pertinent attachments including, but not limited to:

Questionnaire/survey instruments, informed consent, letters of approval from cooperating institutions, copy of external support proposal if applicable.

Yes X No

9. Sources of Subjects: Subjects are United States Air Force Enlisted personnel that will attend the Basic Instructor Course from 15 November 2001 to 7 July 2002. The study will include 111 students with the number of males and females proportionate to the population of the United States Air Force. Enlisted personnel come from differing racial and ethnic backgrounds. All Students used for this experiment will be below the rank of E – 7 (Master Sergeant). Most students will be in their late twenties to early thirties.

10. Study Procedures: Students will be randomly assigned to each class. Twenty classes will be equally divided into a control group and a treatment group. The treatment group will be allowed to use one of the three study skills mentioned in this experiment. The control group will only be able to read the material. The basic design of the experiment is the Post Test Only Control Group design using a $t$ - test for dependent samples. The level of Significance will be set at .05. The Hypotheses to be tested are:

H01: There are no significant differences in the progress check scores between students who use some form of text annotation and students who do not use some form of text annotation.

H02: There are no significant differences, after one week, in the progress check scores between students who use some form of text annotation and students who do not use some form of text annotation.

11. I have gained approval from the Flight Chief of the 82 TRSS, Faculty Development Section. Written consent will not be requested of the individual students in the study to avoid any “demand effects” resulting from the process of gaining consent. It is my belief that the only possible impact of this study is the staff of Faculty Development will be more efficient utilizing limited resources in helping students learn.
12. Confidentiality Safeguards: Students will not be individually identified in the final report nor will they know that the study is in progress. Students will not receive information on the performance of other students.

13. Benefits to Subjects: The results of this study may be used to assist students in managing their study time better. Regardless of the results of the experiments, the data could indicate the best method to follow to get maximum benefit from studying.

14. The Potential risks associated with this study can be overcome with confidentiality of individual subject data in the final report. It is my belief that subjects can only be helped by this study.
MEMORANUM FOR BASIC INSTRUTOR COURSE STUDENTS

FROM: RON BROWN

SUBJECT: CONSENT FOR THE USE OF ACADEMIC DATA

1. I, the undersigned, do hereby give consent allowing the use of my personal academic data specifically relating to an experiment conducted by Mr. Ron Brown. This data pertains to objective 6a through 6f. I understand that this data is being used in conjunction with an experiment on the effects of text annotation on student retention of academic material. I also understand that this data will be used by Mr. Brown to complete requirements for his Doctoral dissertation. If my signature is not given, I understand that my data will be automatically removed from this study.

CLASS NUMBER ___________________

<table>
<thead>
<tr>
<th>NAME</th>
<th>SIGNATURE</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MEMORANDUM FOR: University of North Texas Institutional Review Board of the Protection of Human Subjects 5/4/01

FROM: Mrs. Gayle A. Morrison, Flight Chief, Faculty Development, 82 TRSS

SUBJECT: Approval to use Basic Instructor Course students in study skills experiment.

1. I have spoken with Mr. Ron Brown concerning the use of Basic Instructor Course students in an experiment to observe the effectiveness of certain study skills. I understand that the students will not be aware of the experiment. The students will not be negatively affected in any way and the results of this research may be extremely beneficial to the Faculty Development flight. I approve of his use of the Basic Instructor Course students from 1 May 2002 to 31 November 2002.

Gayle A. Morrison, GS – 13, DAF Flight Chief
TTMS QUESTIONS

7. TTMS stands for:
   e. Technical terminology and Maintenance schedule
   f. Technical Training and Management System
   g. Technical transfer of Mathematical skills
   h. Technical Training and Maintenance System

8. TTMS tracks student information to the:
   e. Course and Class Number level
   f. Course, class, and block level
   g. Class, block, and objective level
   h. Block, objective, and test level

9. TTMS allows the instructor to:
   e. Check future student assignments
   f. Review enlisted force member recruitment contracts
   g. Create schedule instances
   h. Check all students date of rank

10. Instructors have the capabilities to perform all of the following except:
    e. Create an instructor record
    f. Give students credit for SIA
    g. View counseling documentation
    h. Give students credit for a test

11. The software that automates TTMS is known as:
    e. Plato
    f. Plateau
    g. Play-dough
    h. Plethora

12. The Air Force Training Management System (AFTMS) class number is:
    e. A five digit number that comes from your class roster
    f. A Six digit number that comes from your class roster
    g. An alphanumeric code which identifies your class in TTMS
    h. A 15 digit code which identifies your class in TTMS
7. The term which identifies tasks that the student has to complete during course instruction is the:
   - e. Default Status
   - f. PDSCLASS
   - g. Schedule Instance
   - h. Component

21. The term which typically identifies your flight is:
   - e. Domain
   - f. Organization
   - g. PDSCLASS
   - h. Qualification

22. The term which identifies your course and is 15 digits long is the:
   - e. Scheduled instance
   - f. Qualification
   - g. Group instance
   - h. Default status

23. The term that refers to a TTMS generated number which identifies a block of instruction is the:
   - e. Group Instance
   - f. Qualification
   - g. Scheduled Instance
   - h. PDSCLASS

24. The component that you will always choose to find a group of students is the:
   - e. PDSCLASS
   - f. Default Status
   - g. Domain
   - h. Qualification

25. The term organization in TTMS refers to your:
   - e. Flight
   - f. Group
   - g. Wing
   - h. Squadron

26. The term used to identify what a student did in relation to chosen component is the:
   - e. PDSCLASS
27. The term which identifies a class in TTMS and consists of the course number, AFTMS number, shift and section is the:
   e. Qualification
   f. Group Instance number
   g. Schedule instance number
   h. Qualification

28. Which section acts as an human interface between AFTMS and TTMS?
   e. Military Personnel flight
   f. Student Assignments
   g. Registrars Office
   h. Unit Personnel office

29. Who normally creates tracks day to day activities of the students using TTMS?
   e. Instructor
   f. Instructor Supervisor
   g. Course Superintendent
   h. Course Supervisor

30. Who maintains instructor records within TTMS?
   e. Instructor
   f. Instructor Supervisor
   g. Course Superintendent
   h. Course Supervisor

31. In the coming future, Courseware development for automated Instructional Systems Development (ISD) will become available in TTMS with what application software?
   e. Designers Edge
   f. Windows 2001
   g. Office 2001
   h. Microsoft Outlook

32. As the Instructor, you will do all of the following in TTMS except:
   e. Create a Schedule Instance
f. Enroll students into a block
   g. Record block test failures
   h. Assign students to a course

33. **The term which refers to the student social security number is:**
   e. The student ID
   f. The PDSCLASS
   g. The Component
   h. The Domain
BIBLIOGRAPHY


Stordahl, K., Christensen, C. (1956). The Effects of Study Techniques on Comprehension and Retention. Journal of Educational Psychology, 49, 561 569

