THESIS

This is to certify that the

dissertation entitled

THE EFFECTS OF DIRECTIONS TO CHANGE ANSWERS
ON THE ANSWER CHANGING BEHAVIOR
OF COLLEGE STUDENTS

presented by

Patricia Marie Wilson

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Teacher Education

Lois A. Bader
Major professor

Date October 24, 1983

O-12771

THE EFFECTS OF DIRECTIONS TO CHANGE ANSWERS ON THE ANSWER CHANGING BEHAVIOR OF COLLEGE STUDENTS

By

Patricia Marie Wilson

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Teacher Education

1983

ABSTRACT

THE EFFECTS OF DIRECTIONS TO CHANGE ANSWERS ON THE ANSWER CHANGING BEHAVIOR OF COLLEGE STUDENTS

By

Patricia Marie Wilson

This study explored the effect of directions to change answers on the answer changing behavior of college students on an objective final examination for a religion course. Also explored were the differences in the concomitant variables of sex of subjects, Survey of Study Habits and Attitudes (SSHA) scores, and various achievement and ability measures among treatment groups and effectiveness of change groups. The relationship between the level of item difficulty and the effectiveness of answer changing was also investigated.

Procedures

Subjects were 129 students who were enrolled in a liberal arts college in the northeastern United States. Subjects were assigned at random to one of three groups where they were given different sets of directions regarding answer changing on the final examination: no directions, directions to change with caution, and directions to revise to improve scores. Data were analyzed using the analysis of variance and chi-square procedures.

Results and Suggestions

Within the limitations of the methodology and population, the major results of the study were:

1. Directions affected the degree of change effectiveness on easy test items ($p < .01$).

2. The degree of change effectiveness was associated with SSHA work methods scores on difficult items ($p < .05$).

3. Study attitudes and study orientation SSHA scores were associated with the degree of change effectiveness on total items ($p < .05$).

4. Grade point averages ($p < .05$) and final examination scores ($p < .01$) were associated with the degree of change effectiveness on difficult items.

5. SAT mathematics scores and SAT verbal scores were associated with the degree of change effectiveness on difficult items ($p < .01$).

Further investigation into the effects of directions to change answers using different populations and content areas was suggested. Also suggested was further investigation into the reasons students have changed responses and the personality characteristics related to nonchangers and countereffective changers.

To my family,

Dan, Kitty, Sarah, and Kent.

ACKNOWLEDGMENTS

I wish to express my warmest appreciation to Dr. Lois A. Bader, my chairperson and academic advisor. Her wisdom and support have contributed immeasurably to the completion of this doctoral program. I am also grateful to the members of my committee, Dr. James Costar, Dr. Dick Featherstone, and Dr. Eugene Pernell, for their support and flexibility.

I sincerely thank Dr. Lloyd Carr, who contributed his time toward this project, and Dr. Robert Joss and Chris Lamson, who provided excellent dialogue and support throughout this study. Of great help in the planning stages were Dr. Arvo Juola, Dr. Bill Mehrens, and Dr. Tyrus Wessell. They contributed to my thinking and were accessible for dialogue.

The support of my family through this project has been heartwarming. A very special thanks to Mark and Sue Wilson and Byron and Nancy King for entertaining nieces and nephew during summers of coursework and writing. Most of all, my deepest thanks to my husband, Dan, who supported and encouraged me to be a steward of the talents and gifts I have been given.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

THE PROBLEM

Background of the Study

Tests are used widely throughout schools to measure student achievement, to revise curriculum, and to make educational decisions. Millman and Pauk (1969) noted that student test-taking behavior may affect scores and that some students lack sophistication in taking tests. Some students do not make use of the characteristics of the test or the testing situation to improve their scores. They are not test-wise. Millman, Bishop, and Ebel (1965) stated that "it would be desirable to seek ways to reduce differences in test-wiseness among examinees in order to provide more valid estimates of their actual abilities and achievement levels" (p. 724).

According to the taxonomy of test-wiseness proposed by Millman et al. (1965), one aspect of test-wise behavior is going over the questions on a test after one has finished going through the test once and revising the answers that now appear to be incorrect. The effects of such answer revision on objective test scores have been investigated for more than 50 years (Mathews, 1929; McMorris, Lichtstein, & Hoops, 1983). The results have consistently been that students who go back and revise test items tend to raise their scores more often than they lower them (Mueller & Wasser, 1977). Students, however, report that they believe that changing answers will lower

their scores (Foote & Belinky, 1972; Mathews, 1929; Sutton, 1982). Contrary to their stated beliefs, approximately 61 percent to 100 percent of the students revise answers anyway (Mercer, 1978). But not all students revise answers (Stoffer, Davis, & Brown, 1977), and not all students profit from revision. Across the reported studies, the ratio of students who profit from changes to students who lose from changes ranges from 1.4:1 to 48:1 (Hill, 1937; McMorris et al., 1983). Additionally, Ballance (1977) found that only 16.5 percent of his subjects were able to predict their own answer changing behavior.

Study-skills texts and measurement and evaluation texts give mixed advice. Some study-skills texts state that students should stay with their first response (Armstrong, 1975), and others recommend checking answers and revising (Pauk, 1974). Some measurement texts are silent on the matter (Gronlund, 1967; Stanley & Hopkins, 1972), and others encourage teachers "to de-emphasize the conventional caution 'the first answer is invariably the best answer'" (Mehrens & Lehmann, 1978, p. 314), and "those who have studied it suggest that the best strategy is to change one's response if after careful deliberation a better answer or an answer more likely to be correct can be identified" (Hills, 1981, p. 128).

In summary, going back over test responses is an aspect of test-wise behavior. There is mixed advice from the literature in regard to changing responses, and a plurality of students believe that changing answers will tend to lower their scores.

Need for the Study

The conflict between the research findings and student opinion regarding answer changing is not easy to explain. Smith, White, and Coop (1979) posited one theory to explain the discrepancy between research findings and student-reported beliefs. These researchers posited that there might be a relationship between cognitive development and answer revision. Whether or not a student chooses to go back and revise answers might be based on his ability to reverse his logical operations. Because children begin test taking before they are capable of reversible thinking (Piaget, 1972), the response-set in the child might be to stay with his first-choice answer. This response-set might continue to operate.

Reinforcement from negative feedback the student gets when he observes a right answer changed to wrong might encourage the response-set to stay with the first response. Juola (1968) noted that of particular frustration to students are the correct answers changed to incorrect answers. The student remembers that he has lowered his score. He may not even attend to those answers that raised his score. Thus, the tendency from childhood not to reverse is reinforced by the occasional answer revised to a wrong response.

A second theory regarding the discrepancy between student-reported beliefs and actual patterns of change was reported by McMorris and Leonard (1976). These investigators pointed out that as a test-taker goes through an examination his memory may be jogged, and he is more able to reason the correct answer. This viewpoint, they believed, is consistent with the theoretical work of Jenkins

(1974), which suggested that remembering is an active process of elaboration in which stored information is synthesized to form a new construct. That process takes time. The reevaluation of questions on an examination could allow time to pass and thus lead to a more effective reconstruction of the concept required to discern the correct option. But because the reevaluation interrupts the original memory trace, the test-taker may be discouraged from altering his initial response. He does not know which memory to believe, and this conflict may result in avoiding a revision or at least a feeling that his revision might be incorrect.

Recent research has centered on influencing students to change answers and their subsequent answer changing patterns. There have been mixed findings in the studies. The "influences" have varied from reporting to students the findings of the research to rather extensive treatment to alter beliefs (Foote & Belinky, 1972; McMorris et al., 1983; Sutton, 1982).

This study explored the effects of three different sets of directions to change answers on students' answer changing patterns while considering the level of difficulty of the test, the sex difference of subjects, and the achievement and ability of the subjects. Mueller and Wasser (1977) noted that we cannot predict with accuracy which students will gain or lose from changing answers. Thus, giving correct test-wiseness information is not entirely possible from existing research. While some studies have addressed the relationship between sex differences, item difficulty, ability and achievement, and personality and answer changing patterns, no study to date has

investigated the effect of answer changing behavior on self-reported study habits and attitudes. Can a student self-report of study habits and attitudes be used to indicate which students might profit or lose from changing answers?

Statement of Purpose

The purpose of this study was to explore the effects of directions to change responses on objective tests and answer changing patterns and to explore the answer changing patterns associated with the Survey of Study Habits and Attitudes (SSHA) (The Psychological Corporation, 1965). Specifically, the investigator

1. Examined the effect of no specific directions, the effect of directions to revise with caution, and the effect of directions to revise because it will improve scores, on the answer changing patterns of college students who were taking a comprehensive objective final examination for a religion class in which they were all enrolled.

2. Examined the differences in the self-reported study habits and attitudes scores among the direction groups and groups formed by the effectiveness of changing answers.

3. Examined the different patterns of change between the sex of subjects among direction groups and the effectiveness of change groups.

4. Examined the relationship between the level of item difficulty and effective answer changing.

5. Examined the differences in ability and achievement of subjects among direction groups and effectiveness groups.

Research Questions

The major research questions explored were:

1. What is the effect of directions on the number of answers changed?

2. What is the effect of the directions on the percentage of correct changes?

3. Are there differences in SSHA scores among the direction groups (treatment) and the effectiveness groups (countereffective, ineffective, and effective changers)?

4. Is there a difference between easy and difficult item changes and the effectiveness of the change?

5. Are there any differences in the percentage of correct changes among the direction groups and the groups formed by sex of subjects?

6. Are there any differences in the achievement measures among the direction groups and the effectiveness groups?

7. Are there any differences in the ability measures among the direction groups and the effectiveness groups?

Generalizability

The population was one group of 129 students who were enrolled in a required freshman-level religion class in a private four-year liberal arts college north of Boston, Massachusetts, in the fall of 1982. The subjects were primarily of the Caucasian race and were primarily from middle-class homes. The mean SAT score for the population was 485 and the mean SSHA score was 101, which is approximately the 30th percentile. There were 51 males and 78 females in the

population. Most of the subjects were in the first quarter of their freshman year. The course test used to collect data was constructed by the course instructor. It may be concluded that the findings in this study can be generalized to populations in other colleges with similar enrollments.

Limitations

The limitations of this study were the level of difficulty of the course test, the length of the course test, and the subsequent willingness of the students to revise answers when they felt pressed for time. Also, the willingness of the students to read the directions limited the findings. An additional limitation was the self-reporting format of the SSHA.

Definition of Terms

The reader will better understand this study if certain terms are clarified.

Answer changing or revision: Any observed blot out, erasure, or crossed-out answer.

Answer changing behavior or pattern: The number, rate, or frequency of changes and/or the profit or loss from the change.

Directions to change: The printed directions to subjects regarding changing or revising answers.

Direction of change: Profit or loss from a change.

Net-gain changes: The gain from changes, which is calculated by subtracting the right-to-wrong changes from the wrong-to-right changes.

Degree of change effectiveness: A categorization of change effectiveness based on a percentage of correct changes; 0-39 percent = counter-effective, 40-59 percent = ineffective, 60-100 percent = effective.

W-R: Wrong-to-right changes.

R-W: Right-to-wrong changes.

W-W: Wrong-to-wrong changes.

Item difficulty: The percentage of the total group passing the item.

Difficult item: Those items where 30 percent or more of the students missed the item. The mean item difficulty was 29.

Easy item: Those items where 29 percent or fewer of the students missed the item.

Total items: The total number of items changed as opposed to easy items changed and difficult items changed.

Index of discrimination: Calculated by subtracting the percentage of the upper group marking the correct answer and the percentage of the lower group marking the correct answer.

Discriminating item: Items whose index of discrimination is 23 percent or above. The mean item discrimination was 23.

Test-wise: The ability to use the format or characteristics of a test or testing situation to receive a higher score.

Effective changing: Changing 60-100 percent of answers from wrong to right.

Ineffective changing: Changing 40-59 percent of answers from wrong to right.

Countereffective changing: Changing 60-100 percent of answers from right to wrong.

Penalty-for-guessing scoring directions: Wrong answers are subtracted from right answers; also called correction-for-guessing scoring directions.

Organization of Subsequent Chapters

The content of Chapter I included background of the study, need for the study, the research questions, the generalizability and limitations of the study, the definition of terms, and organization of subsequent chapters. In Chapter II the research and literature related to the study are reviewed. A description of the design and methodology used in the study is contained in Chapter III. In Chapter IV, the data are presented and analyzed. In Chapter V, a summary of the study, discussion of results, and appropriate conclusions and recommendations for future research are presented.

CHAPTER II

REVIEW OF RELATED LITERATURE

Introduction

The purpose of this study was to investigate the effects of directions to change answers on the answer changing behavior of college students. Also explored were the differences in the concomitant variables of sex of subjects, Survey of Study Habits and Attitudes (SSHA) scores, and various achievement and ability measures among treatment groups and effectiveness of change groups. The relationship between level of item difficulty and the effectiveness of answer changing was also investigated.

The review of the literature is organized under the following headings: (1) Beliefs About Answer Changing; (2) Subjects, Test Content, and Changes; (3) Test Directions; (4) Achievement and Aptitude, (5) Item Characteristics, (6) Sex Differences, and (7) Personality.

Beliefs About Answer Changing

Beliefs concerning the advisability of changing answers on objective tests have been reported in empirical research, in textbooks on measurement and evaluation, in journal articles, and in study-skills textbooks. The empirical research is reviewed first.

In a survey, Mathews (1929) found that 86 percent of his respondents felt that revising answers would lower their scores. Eleven percent felt changing answers would raise scores, and 13 percent felt that changes would cancel each other out. In an informal show of hands, Lynch and Smith (1975) found that 75 percent of the 300 subjects believed that changes would lower their scores. Jacobs (1972) found 30 percent of his subjects believed changes would result in a gain, 45 percent said that changes would produce a loss, and 25 percent said they did not know. Jacobs found no relationship between the subjects' reported beliefs and mean net gains from their answer changing.

Foote and Belinky (1972) assessed student beliefs concerning answer changing before attempting to influence them to revise answers. Ninety-three percent of their subjects responded that they did check over their answers after tests, and 99 percent said they had heard first-impression answers were more apt to be correct. Thirty-two percent believed revisions would improve scores, and 64 percent believed revisions would lower scores or make no difference. After examining their own answer changing on a returned examination, 72 percent of the students responded that they were surprised at the results. However, only 48 percent of the students said they would change answers more often.

Mueller and Shwedel (1975) found no one in their sample who believed their answers would be improved with revisions. Stoffer, Davis, and Brown (1977) found that 80 percent of their subjects believed changes would lower scores. Smith, White, and Coop (1979)

found only 7 percent of their sample who believed that answer revision would increase scores. Eighty-six percent of their sample, however, did revise one or more answers. Similar beliefs and revision patterns were noted by Foote and Belinky (1972), Jacobs (1972), and Mueller and Shwedel (1975).

Ballance (1977) studied student expectations concerning answer changing and student motivation to change answers. Findings were that student expectations regarding the outcomes of their revisions were not related to the actual outcomes. Only 16.5 percent of the students were able to predict their own behavior. Also found was that 86 percent of the students would change answers because of reassessment of the question. McMorris (1981), Sutton (1982), and McMorris, Lichtstein, and Hoops (1983) also found no relationship between self-reported student beliefs and answer changing practices. McMorris (1981) found a mean of 3.5 on a seven-point scale that attempted to measure the advisability of changing answers. Students reported they felt that changing responses was "somewhat advisable." This finding was replicated by McMorris et al. (1983). McMorris (1981) also investigated the reasons for which students would change responses. From 65 to 84 percent of the students reported they would change responses if they needed to correct clerical errors, if they better understood the question, and if they had rethought the question. The relationship between these student reports was compared to student gains. There was a tendency for those with a more positive attitude toward changing to gain more, but the gains were not significant. McMorris et al. (1983) did find attitudes toward

changing answers on the course test (but not on other tests) significantly related to the percentage of changes and gains. Their subjects were instructed on the benefits of change. The strongest predictor of change, however, was the number of answers changed.

Sutton (1982) collected data on student beliefs concerning answer changing, established control and experimental groups, and attempted to alter the subjects' beliefs concerning answer changing. Results were that posttest measures were significantly higher for the experimental group for beliefs concerning the value of answer changing, the frequency of answer changing, and the net gains from answer changing. Sutton also found that students profited more from revisions due to clerical errors than from changes of items that had been rethought.

Sutton surveyed 11 different current measurement texts to assess their comments concerning answer changing. He found that Mehrens and Lehmann (1978) addressed the issue of answer changing in a chapter on assembling and analyzing classroom tests. Mehrens and Lehmann pointed out that the empirical evidence supports the view that changing answers most often benefits the test-taker. Ebel (1972) also stated that students should review answers and correct any errors. Sutton did not locate other texts that included information about answer changing.

The present investigator perused five current texts and found one, in addition to Mehrens and Lehmann (1978) and Ebel (1972), that advised students to go back over and catch careless errors and change responses after careful deliberation and also reported the

research evidence concerning answer changing. This information was given in a section of the text devoted to developing test-taking skills (Hill, 1981).

Millman, Bishop, and Ebel (1965) included answer-changing behavior in their taxonomy of test-wiseness. These authors recommended that students use time remaining after completion of the test to reconsider answers. "Change answers if it seems desirable. Examinees generally increase their scores if they do" (p. 714). Further, Millman et al. stated that it is desirable to reduce the differences in test-wiseness among examinees in order to provide more solid estimates of actual abilities and knowledge. "How can we change test items and test directions, or other conditions of test-administration, to minimize the harmful effects of differences in test-wiseness?" (p. 724). However, in a 1979 review of literature on test-wiseness, Sarnacki did not include a review of answer-changing research.

Finally, current study-skills texts were reviewed. Lynch and Smith (1975) reviewed several texts that instruct students to stay with their first answers. This investigator surveyed some later editions of study-skills texts. Armstrong (1975) instructed students not to change answers because one's first response is more reliable. Maiorana (1980) followed the same general thought and advised students to avoid impulse changes. Staton (1982) advised, "Do not change your answer because of a mere doubt" (p. 69).

Shepherd (1982) recommended that students exercise care in changing responses because "some students have a tendency to change correct answers to incorrect answers" (p. 18). He further recommended

that students take note of their own behavior regarding changing responses until they establish their own profit-or-loss patterns. Gilbert (1982) stated that students have only three chances in ten to change from the incorrect answer to the correct answer. Both Pauk (1974) and Robinson (1970), however, recommended that students change answers if they believe that another answer is correct.

In summary, the empirical research concerning answer changing suggested that the majority of students report that they believe answer changing will decrease their scores, but they revise their answers anyway. Mixed results have been found when researchers have attempted to influence answer changing behavior. Many current texts on measurement and evaluation do not include sections on developing test-taking skills and do not include information regarding answer-changing behavior. Major reviewers of research on test-wiseness have not agreed on the inclusion of answer changing in their works. Finally, study-skills texts give mixed advice concerning changing answers.

Subjects, Test Content, and Changes

Beginning with Lehman (1928) and Mathews (1929), the majority of studies concerning answer changing have been conducted with subjects who were college undergraduates or graduates. Further, most of the studies were conducted with college students who were enrolled in education or psychology courses in which objective tests constructed by the professors were used to collect data on answer changing patterns (Archer & Pippert, 1962; Ballance, 1977; Bath, 1967; Berrien, 1939; Clark, 1962; Foote & Belinky, 1972; Greene, 1981; Hill, 1937; Jacobs,

1972; Jarrett, 1947; Lamson, 1935; Lowe & Crawford, 1929; Lynch & Smith, 1975; McMorris et al., 1983; McMorris & Leonard, 1976; Pascale, 1974; Penfield & Mercer, 1980; Range, Anderson, & Wesley, 1982; Reile & Briggs, 1952; Smith et al., 1979; Sitton, Adams, & Anderson, 1980; Sutton, 1982). Vidler and Hansen (1980) used college psychology students as their subjects, but they used the Watson-Glasser Critical Thinking Test as their data-gathering instrument. Stoffer et al. (1977) examined course examinations of college psychology students and course examinations of Air Force personnel who were enrolled in a technical-vocational skills course.

A few researchers have studied answer changing behavior of college students who were not enrolled in education or psychology courses. Mallinson and Miller (1956) studied answer changing behavior on the objective examinations of students enrolled in liberal arts courses. They collected data from 1,029 college students on tests from courses in biology, economics, political science, and chemistry, in addition to psychology and education. Copeland (1972) examined the final examinations of chemistry students. Reiling and Taylor (1972) examined economics tests, and Ballance and Gentzel (1980) looked at junior-college students' business-principles examinations. Davis (1975) examined tests for microbiology, pharmacology, physiology, and pathology courses that were taught in a medical school. Ballance (1975) examined the tests of 144 first-year professional students from the allied-health-care field.

Two studies examined the General Education Development Tests for answer changes (Smith, 1975; Smith & Moore, 1976). This

test contains subtests for English usage, social studies, natural science, literature, and mathematics. These studies extended the research beyond the college-age population to a noncollege adult population.

The research was extended to the middle- and elementary-school population in the late 1970s. Crocker and Benson (1980) examined the mathematics-computation subtest of the Metropolitan Achievement Test. His subjects were in grade seven. Using subjects from grades seven and eight, Cummings (1981) examined the Nelson Reading Skills Test for answer changing. In New Zealand, Jackson (1978) examined the Progressive Achievement Tests: Study Skills for response alterations of 10-, 12-, and 14-year-old subjects. Four subtests of the Iowa Test of Basic Skills for pupils in grades four, five, six, and seven were examined by Mercer (1978). This study involved the first use of a predominantly nonwhite population. Subjects in this study were also grade-level and non-grade-level examinees who were administered the test on grade level or below grade level, according to teacher judgment.

Payne (1983) used science-course tests of subjects in grade eight. Her sample included nonwhite subjects. Both Mercer (1978) and Payne (1983) drew their samples from urban schools. Beck (1978) extended the research to subjects in grade three who took the Metropolitan Reading Test. In most of these studies, the format of the objective tests was four-option multiple-choice items. However, Berrien (1939), Hill (1937), Jarrett (1948), Lamson (1935), Lehman

(1928), Lowe and Crawford (1929), Lynch and Smith (1975), Mathews (1929), and Mueller and Shwedel (1975) examined changes on true-false items.

Whether standardized or teacher-made examinations; true-false or multiple-choice; college, adult, middle-school, or elementary-school subjects, the investigators all found that when the changes were tallied more answers had been changed from wrong-to-right than from right-to-wrong (Mueller & Wasser, 1977). Mercer (1978) reported a range of 40 percent to 73 percent of answers changed from wrong to right on studies conducted through 1978. This range has been replicated in studies to the present (Ballance, 1975, 1977; Beck, 1978; Best, 1979; Crocker & Benson, 1980; Cummings, 1981; Greene, 1981; Jackson, 1978; McMorris, 1981; McMorris et al., 1983; Mercer, 1978; Payne, 1983; Penfield & Mercer, 1980; Range et al., 1982; Smith et al., 1979; Stoffer et al., 1977; Sitton et al., 1980; Sutton, 1982; Vidler & Hanson, 1980).

In a review of eight studies, Mueller and Wasser (1977) reported a median gain-to-loss ratio of 2.7:1 items. Mercer (1978), summarizing the data from 23 studies, reported a range of .54 to 23.4 mean net gains for subjects. McMorris et al. (1983) reported a student gain-to-loss ratio of 48:1. The range of student gains appeared to go from this high down to a student gain-to-loss ratio of 1.4:1 reported by Hill (1937).

In summary, the majority of the studies concerning answer changing have been conducted using subjects who are college students enrolled in education or psychology classes. A few studies have

investigated answer changing on examinations from science courses and liberal arts courses. Recently, the research has been extended to younger students of middle school and elementary age. These studies have used standardized tests as the data-gathering instruments. The research has also been extended to the adult non-college-student population. In all studies, results have indicated that more changes are from wrong-to-right than from right-to-wrong.

Test Directions

In most of the studies on answer changing, the tests were administered without specific directions regarding answer changing, and the subjects were not aware that an experiment was being conducted. The researchers collected the tests of the students and examined them for erasures or crossed-out answers and then tallied the data into categories of wrong-to-right changes (W-R), right-to-wrong changes (R-W), and wrong-to-wrong changes (W-W). There were, however, several studies in which special answer sheets or special directions regarding changes were given.

Mathews (1929), who collected data from a true-false, multiple-choice, objective examination, told students that changes were more apt to be correct than incorrect. He also included penalty-for-guessing scoring directions (wrong answers subtracted from right answers) on the true-false section of the test. On the multiple-choice section, 2.5 percent of the answers were changed, and 3.16 percent of the answers on the true-false section were changed. Almost 60 percent of the true-false items were changed to correct, whereas only 50

percent of the multiple-choice items were changed to correct. True-false questions had 56 (9.8 percent) of the answers omitted, whereas there were 5 (.9 percent) omitted multiple-choice answers. Mathews concluded that the higher true-false omission rate influenced the more positive change rate for true-false items.

Hill (1937) and Ballance and Gentzel (1980) also included a penalty-for-guessing scoring direction. However, they did not inform their students of a benefit for answer changers. Hill (1937) reported a lower change rate and score improvement for changing than previous researchers. There was a 1.43:1 ratio of correct to incorrect changes. However, 2.3 percent of the answers to questions were omitted. Ballance and Gentzel (1980) studied answer changing under three different sets of scoring directions. The No Penalty group (N = 29) was told that scoring would be based on the number of questions they answered correctly. The Correction for Guessing group (N = 36) was told that scoring would be based on the number correct minus one point for each wrong response. The Severe Penalty group (N = 40) was told that scoring would be based on the number right minus the number wrong. There was a significant difference in the mean changes from treatment to treatment. There was, however, no benefit for one group over another in terms of profits from changes. The authors concluded that correction-for-guessing directions appear to contribute to "a testing atmosphere in which better decisions are made about answer changing" (p. 654).

A number of researchers have included some form of directions to change answers in their studies. Lehman (1928) studied the answer-changing behavior of students by using a two-column answer sheet. Lynch and Smith (1975) provided a similar sheet. Lowe and Crawford (1929) developed a rotation experiment to assess the differences between two groups of subjects who took two parts of a test using two different procedures; both groups used one procedure on the first part of the test and then switched procedures for the second section of the test. One procedure involved reading through the test first and waiting to mark items until the second reading. The other procedure involved reading and marking choices and then reading and revising. Significant gains were noted for those who read and marked and then revised. In addition, 10 percent of the test items were revised under the directions to read and mark and then revise.

Jacobs (1972) conducted an experiment to attempt to measure changes from first impressions to final answer. He projected the questions via slides for 30 seconds. Subjects were to record their answers quickly and did not know they would have an opportunity to revise their work. Then subjects were given a mimeographed version of the projected questions and allowed to revise using a different-colored pencil. The revision rate was 32 percent. The mean number of changes was 14.7, and the ratio of points gained to points lost was 2.8. The results of this study are limited by the unique data-gathering technique of first-impression answers.

Foote and Belinky (1972) gave information and feedback concerning answer changing to two classes of psychology students who had

already taken two course examinations on which their answer changing patterns had been analyzed. Results showed that subjects' answer changing patterns on the last two course examinations were not significantly different from the patterns on the first two examinations in regard to either the rate or effectiveness of changing. On a questionnaire, only 48 percent of these students had indicated that their future answer changing rate would increase.

Based on the no-change findings of Foote and Belinky, Sutton (1982) attempted to influence student answer changing by influencing and changing student beliefs. Subjects were 184 students in a teacher-education program conducted by the Air Force. The subjects were two intact groups that attended classes during the first and then second five-week time period in a ten-week session. Using small-group instruction, the experimental group was informed about the benefits of answer changing and given individual feedback concerning their own changes on tests. Pretreatment equality was established between groups on beliefs concerning answer changing, frequency of changing answers, and gains from answer changing. Results at the end of the treatment showed significant differences between the control and experimental groups for beliefs, rate of changes, and effectiveness of changes. The mean net gain score for the control group was 1.61, and for the experimental group, 2.3.

Using a sample that had been informed of the benefits of answer changing, McMorris et al. (1983) investigated the relationship between the attitude toward and knowledge concerning answer changing and the patterns of answer changing. Results indicated that students

who were more positive toward answer changing and to the course and its test were more likely to benefit from changes. The percentage gain was a mean of 3.1, and the student gain-to-loss ratio was 48:1. Still, the most significant predictor of gain was the number of changes made. This study was not designed to measure the effect of influence to change answers on answer changing. The results were limited by the design.

In summary, a few studies included directions to change answers in their designs. Some of these studies also included a penalty-for-guessing scoring instruction. Under penalty-for-guessing scoring instructions, fewer answers were revised and more answers were omitted. There were inconsistent findings in studies where there was an attempt to influence student answer changing through instructions to change answers or feedback concerning student answer changing. Some of these studies found a higher rate of change, some found no difference in the rate of change, and some found profits from changes.

Achievement and Aptitude

Several studies have investigated whether high-scoring subjects or low-scoring subjects make more or fewer answer revisions and whether those revisions tend to result in a profit or loss for them. In addition, a few studies have investigated the relationship between subject grade-point averages and answer changing, subject course grade and answer changing, and subject ability and answer changing.

In a number of studies, high-scoring students made more gains from changing answers than did lower-scoring students (Archer & Pippert, 1962; Ballance, 1975; Bath, 1967; Berrien, 1939; Best, 1979; Clark, 1962; Copeland, 1972; Crocker & Benson, 1980; Davis, 1978; Jarrett, 1947; Lehman, 1928; Lynch & Smith, 1975; Mallinson & Miller, 1956; Mathews, 1929; McMorris, 1981; Payne, 1983; Reile & Briggs, 1952; Reiling & Taylor, 1972; Sitton et al., 1980; Smith, 1975, 1977; Stoffer et al., 1977). It might be noted that in most of these studies the total test score was used to investigate the net gains, and this score included the net gains. Since answer changing tends to raise the total test scores, there is a reciprocal causal effect (Mueller & Wasser, 1977).

Pascale (1974), Penfield and Mercer (1980), and Smith et al. (1979) did not find gains for high-scoring subjects. Penfield and Mercer (1980) found that high-scoring students changed from a wrong alternative to another wrong alternative more often. Mallinson and Miller (1956) noted a similar finding, although their high-scoring subjects maintained top position for improvement. Smith et al. (1979) found that lower and average students profited more from change.

Archer and Pippert (1962), Lehman (1928), Lynch and Smith (1975), and Mathews (1929) found that lower-scoring students had the highest rate of changes. There was, however, an inverse relationship between the number of changes and the gains. Johnson (1978) also noted a higher rate of changes among lower-scoring students.

McMorris et al. (1983), Mueller and Shwedel (1975), and Reiling and Taylor (1972) reported a positive relationship between

the number of answers changed and the gains from changes. McMorris et al. noted that the most highly significant predictor of gain was the number of answers that were changed (r = .77)

Several authors reported that high-scoring students made fewer changes than low-scoring students (Best, 1979; Johnston, 1978; Lehman, 1928; Lynch & Smith, 1975; Mathews, 1929; Mueller & Shwedel, 1975; Reile & Briggs, 1952; Sitton et al., 1980; Smith & Moore, 1976). Bath (1967), Berrien (1939), and Mercer (1978) found no differences between high- and low-scoring subjects. In all these studies, there were inconsistencies across groups.

Jacobs (1972) found no relationship between verbal intelligence, as measured by the Quick Word Test, Level II and net gains. Stoffer et al. (1977) found no relationship between net gains and the verbal section of the Scholastic Aptitude Test for their college group and no relationship between the Airman Qualifying Examination and net gains for their technical-skills sample. McMorris and Leonard (1976) found no relationship between student-reported grade-point averages and net gains. This finding was replicated by Ballance (1975), and McMorris et al. (1983).

There have not been consistent findings in regard to the relationship of course grades and profits from changes. Reiling and Taylor (1972) found no relationship between the final course grade and answer changing, although the poorer students were slightly better answer changers. Bath (1967) and Reile and Briggs (1952) found the top and bottom quarters of subjects to profit more from changes. Sitton et al. (1980) found that there was a relationship

between A and C students and successful answer changing. Range et al. (1982) noted that B students made the greatest gains from answer changing.

In summary, there have been inconsistent results concerning the relationship between answer changing behavior and measures of achievement and ability. Several studies have reported a relationship between high scoring students and effective answer changing, whereas others have reported a relationship between low scoring students and high change rates. Studies to date have not reported a significant relationship between answer changing behavior and grade point averages. Additionally, the SAT scores have not been found to be related to answer changing patterns.

Item and Test Characteristics

Studies concerning item difficulty have had inconsistent findings. The majority of the studies found that difficult items were changed more often than easier ones (Beck, 1978; Greene, 1981; Jackson, 1978; Lynch & Smith, 1975; McMorris et al., 1983; Vidler & Hansen, 1980). Studies by Best (1979) and Jacobs (1972) found that items of moderate difficulty and easy items were changed most often. Ballance (1975) found that there were no differences in the rate of or profit from changes of difficult items. Jacobs (1972) and Vidler and Hansen (1980) found that difficult items were changed with less success.

Reiling and Taylor (1972) and Smith et al. (1979) investigated the patterns of changes on higher cognitive questions and lower

cognitive questions. Findings were that higher cognitive questions were changed more often, but there was no relationship with net gains or losses.

Two studies investigated the position of items in the test and the frequency of changes. The findings were conflicting. Jackson (1978), studying the behavior of late-elementary-school children, found that more items were changed at the end of the test and at the beginning of sections within the subtest. Jackson's data-gathering instrument was a standardized achievement test. Reile and Briggs (1952), studying college students, found fewer changes were made on items near the end of the test.

Crocker and Benson (1980) investigated test reliability and item changing. They found the internal consistency of the test was virtually unaffected by examinees' response changes. They noted a slight decrease in the standard error of measurement, but they felt the decrease was too small to have practical significance. Cummings (1981) replicated these findings. He noted that item-discrimination power was reduced by changes, whereas Crocker and Benson (1980) found that item-discrimination power was increased by changes.

In summary, there have been inconsistent findings regarding item and test characteristics. Some studies have reported significant gains and rate changes for easy items, and other studies have reported significant gains and rate changes for difficult items. One study reported more changes on beginning test items, and another reported more changes on ending test items. In addition, there have been inconsistent findings concerning item discrimination. Finally, two

studies that investigated test characteristics reported that the internal consistency of the test was not changed due to answer revisions.

Sex Differences

Authors who investigated sex differences reported inconsistent findings. Bath (1967), Copeland (1972), and Stoffer et al. (1977) all found that females benefited more from changes than did males. Copeland found that males made more changes, whereas Reile and Briggs (1952) had found that females made more changes but the males had more benefits from their fewer changes. Mueller and Shwedel (1975) found that males made more changes, but they were right-to-wrong changes. Mercer (1978) found that seventh- and eighth-grade males made significantly more changes on the mathematics subtest of the Iowa Test of Basic Skills. Pascale (1974) and Reiling and Taylor (1972) found that males benefited from changes more often than did females. Beck (1978) and Sitton et al. (1980) found no differences in the number of changes or gains between males. Mueller and Wasser (1977) noted that throughout the studies they reviewed, the sex with the highest test mean had more profit from change.

Personality

Ballance (1977) and Lynch and Smith (1975) suggested that the relationship between personality and answer changing be explored. From the mid-1970s to the present, a few studies have attempted to explore that relationship (Greene, 1981; McMorris & Leonard, 1976; Payne, 1983; Range et al., 1982; Sitton et al., 1980; Stoffer et al.,

1977). For the most part, these studies have used varying personality measures. The most studied variable has been anxiety.

McMorris and Leonard (1976), using students in four educational-measurement and evaluation courses who were studied as four groups, explored the relationship between item-response changes and four cognitive styles: impulsivity-reflectivity, anxiety, preference for complexity, and field independence. In this study, anxiety was measured by Spielburger, Gorsuch, and Lusheme's Self-Evaluation Questionnaire (1968). Results indicated that most of the relationships were quite low and that the relationships were not consistent across the four groups. Of the variables studied, impulsivity was the characteristic most related to answer changing and the one the authors felt might show replication in further studies. There did not appear to be a relationship between anxiety and answer changing.

Stoffer et al. (1977) found that 29 percent of their college psychology students and 20 percent of their military technical-vocational subjects made no answer changes on most of the course tests during the semester. The observation was true across several tests, suggesting to the investigators a style of test taking as a "changer" or a "nonchanger." They posited that those who perceived that their test scores (reinforcements) were not under their control would change fewer answers than those who perceived that their test scores were under their control. Findings, however, were that there was no correlation ($\underline{r}$ = .00) between Rotter's personality measure for internal-external locus of control and answer changing.

Sitton et al. (1980) studied undergraduate students enrolled in adolescent psychology. They explored the relationship between answer changing and depression, anxiety, assertiveness, introversion-extroversion, and several demographic variables. The Taylor Manifest Anxiety Scale was used to measure anxiety. Results were that assertive students and top-scoring students changed fewer answers. Using a multiple-regression procedure, the four best predictors of the number of changes were the LIE scale of the Eysenck Personality Inventory, assertiveness, course grade, and depression. These predictors accounted for 22 percent of the variance in answer changing. The three best predictors of wrong-to-right changes were the LIE scale, marital status, and assertiveness. The LIE scale, according to the inventory's authors, "may be used to identify subjects showing a 'desirability response-set'" (Eysenck & Eysenck, 1968). In discussion, Sitton et al. suggested that the connection between a desirability response set and patterns of answer changing might include a motivation-level difference or a difference in ego investment in academic achievement and thus represent more of a desire to respond to a socially desirable pattern. Subsequent research by McMorris et al. (1983) found a significant relationship between student attitude toward answer changing and the number of answers changed and the number of answers changed wrong to right. The investigators found no relationship between anxiety and answer changing.

Greene (1981), Payne (1983), and Range et al. (1982) continued to look for a relationship between anxiety and answer changing. Greene investigated the answer changing behavior of graduate subjects

enrolled in a statistics course and found that high-anxious students made more changes than did low-anxious students. Both the high-anxious and low-anxious students profited equally from their changes. High-anxious students had a slight tendency to reduce their scores by their changes. Greene used the Test Anxiety Scale to measure anxiety.

Range et al. also found that high-anxious students changed answers more often. The State-Trait Anxiety Scale (Spielberger) was used to measure anxiety. In addition, these authors investigated the relationship between depression, sex-role stereotyping, and self-actualization and answer-changing patterns. In addition to the results concerning anxiety, findings were that nondepressed students made more positive changes and that the best predictors of these positive changes were compulsivity, a positive view of man, masculinity (sex-role typing), lack of depression, and anxiety. These accounted for 33 percent of the variance.

Studying younger students in grade eight, Payne (1983) replicated the findings of Green (1981) and of Range et al. (1982) in regard to anxiety and changes. Payne found that high-anxious subjects changed more answers. The anxiety measure in the study was the Survey of Feelings about Tests by Harnish, Hill, and Lyons (1980). Payne also found that black subjects were more anxious than white subjects, that their mean test scores were much lower, and that they changed significantly more answers wrong to wrong and right to wrong than did the white subjects. However, there was no correlation between anxiety and answer changing for the group of blacks. The findings

of this study were consistent with Ebel's (1965) report that there is a negative correlation between level of achievement and level of test anxiety.

In summary, McMorris and Leonard (1976) and Sitton et al. (1980) were unable to find a strong relationship between anxiety and answer changing patterns. Their findings were in conflict with those of Greene (1981), Payne (1983), and Range et al. (1982), who all found that high-anxious students change more answers. McMorris and Leonard found indications that impulsivity might have a relationship with answer changing, and Sitton et al. found a relationship between the LIE scale of the Eysenck Personality Inventory, which is thought to measure desirability response set, and patterns of answer changing. In addition, Range et al. found that nondepressed students made more positive changes.

Summary

The related literature that was reviewed in this chapter was concerned with answer changing on objective tests. More specifically, studies were reviewed that focused on answer changing and beliefs concerning answer changing, item and test characteristics and answer changing, achievement and answer changing, ability and answer changing, sex differences and answer changing, and personality and answer changing. On the basis of this information, those factors that appear to be related to answer revision were selected for the present study.

The review of the literature suggested a need to explore the effects of test directions to revise answers under experimental

conditions. A need to evaluate the differences between groups given varying directions and groups formed by change effectiveness and the variables of sex, item difficulty, achievement, and ability was also suggested. Finally, a need to explore the differences in study habits and attitudes scores and answer changing behavior under directions to change answers was also implied by the existing research.

Chapter III presents the population, methodology, and data-gathering procedures, as well as the design, hypotheses, and statistical treatment of the data.

CHAPTER III

METHODOLOGY

Introduction

This chapter identifies and describes the population and elaborates on the methodology used in conducting the study. The data-gathering procedures are outlined, the design of the study is presented, the data-gathering instruments are described, and the hypotheses are presented. Finally, the statistical treatment of the data is explained.

Population

The subjects in this study were college students who were enrolled in a 100-level religion class at a small liberal arts college north of Boston, Massachusetts. The students who attend the college are from middle- or upper-middle-class backgrounds. The majority of the students are from the northeastern United States. There were 159 students who originally enrolled in the religion course. Due to attrition, 30 were eliminated from the study. Twelve students dropped the course. Data were unavailable for 18 other subjects.

The religion course was selected because of the number in the class and the age of the students enrolled. The course is a core requirement for all students. The majority of the students were entering freshmen. Thus, the data for this study were collected

on freshman students who were in their first three months of college life.

Procedures

Permission for this study was obtained from the academic dean of the liberal arts college. The registrar granted permission to obtain the needed Scholastic Aptitude Test (SAT) scores and grade point averages. The instructor of the course volunteered to allow the investigator to obtain the data for the study.

The Survey of Study Habits and Attitudes (SSHA) by Brown and Holtzman (Psychological Corporation, 1965) was administered by the investigator to the subjects in a 35-minute block of class time the week before the final examination. Permission from the college's retention committee was obtained to tell the subjects that the survey was to be used to study retention of students at the college. Subjects were also told that the results of the survey would not influence their grades and that their scores on the survey would be discussed with them if they requested a conference.

Approximately 40 of the subjects were absent during the class period when the survey was administered. These 40 took the survey at the time of the final examination. A cover letter explaining the purpose of the survey was given to these students. (See Appendix A). The letter explained that the survey was being administered for retention purposes and gave directions for taking the test. These materials were distributed at the door with the final examination. Most of the students received a personal word concerning the survey.

Subjects were assigned at random to one of three groups: control, treatment 1, or treatment 2. Each of these groups received a different set of printed directions on the objective section of the comprehensive final examination. The control-group subjects received the following directions: "Please read the instructions carefully and answer the questions asked. If you make an error, cross out (X) the wrong answer and write another answer." The students in treatment 1 were instructed, "If you feel that you have made an error, cross out (X) the answer you believe is wrong and write another answer. Change with caution answers you believe are incorrect." Students in treatment 2 received these instructions: "If you find that you have made an error, cross out (X) the answer you believe is wrong and mark another answer. After you have worked through the test once, go back and change the answers you now believe are incorrect. Research indicates that revising answers will improve your score on this test." In addition, treatment 2 students received the following direction at the conclusion of the examination: "GO BACK AND REVISE YOUR ANSWERS!!" Each set of instructions was headed with "INSTRUCTIONS," and an attempt was made in the printing to space and line so that the instructions appeared to be of equal length. (See Appendix B.) Additionally, the objective section of the final was placed in the middle of the test packet. A sentence-completion section preceded the treatment portion, and an essay section followed.

The examinations were arranged in alphabetical order. Two teacher assistants and the investigator distributed the examination from outside the examining room. The investigator's presence was

explained by the cover letter of the survey. The teacher assistants and the investigator attempted a word of explanation to each student. Further, because the investigator had spoken in classes and taught approximately 50 of the subjects in two sections of study skills, the subjects did not question her presence.

The SSHA was collected with the final examination and then separated from the finals, which were graded by the instructor of the religion course. When the investigator examined the objective section of the final for changes, a tally was made of the answers that had been changed from right-to-wrong (R-W) and wrong-to-right (W-R). A wrong-to-wrong category was eliminated because 8 percent of the changes had been thoroughly blotted out. Items that were erased, blotted out, and crossed out were tallied as changes. The objective section of the final examination was then transferred from the printed-test answer blanks to machine-scored answer sheets. The answer sheets were then processed at the Testing Center of the Michigan State University Computer Center. The examinations were scored; an item analysis was computed; and the mean, standard deviation, variance, standard error of measurement, and reliability (Kuder-Richardson formula #20) were computed.

The Scholastic Aptitude Test (SAT) scores were obtained for all subjects, as were the end-of-the-quarter or cumulative grade point averages. The study was conducted with 129 subjects. There were 45 in the control group, 40 in the treatment 1 group, and 44 in the treatment 2 group. Attrition in the population was due to 12 who dropped the course or failed to take the final examination and 18 for

whom SAT or SSHA scores were not available. SAT scores were available for all freshman students, but the SAT scores are not required by the college for transfer students. Therefore, the majority of subjects who were dropped from the study were transfer students.

Design

The study employed the randomized control-group posttest-only design. The 159 subjects were assigned at random to one of three groups: control, treatment 1, or treatment 2.

The test directions were the treatment or independent variable, and the posttest or dependent variable was the answer changes. The intention behind the use of the SSHA scores was to discover whether any study habits or attitudes are associated with effective or ineffective answer changing. The associations between answer changing effectiveness and sex, item difficulty, objective test scores, grade point averages, and SAT scores were also explored.

Instruments for Data Gathering

Two instruments for gathering data were used. They were the Survey of Study Habits and Attitudes (SSHA) by Brown and Holtzman (Psychological Corporation, 1965) and a comprehensive final multiple-choice and true-false test that was constructed by the professor of the religion course.

The SSHA is a 100-item survey of study habits and attitudes for grades 12 through 14, which is a report of self-perceived study habits and attitudes. The manual (1967) recommends its use as a screening instrument, a diagnostic instrument, a teaching aid, and a

research tool (pp. 5-6). The administration time is approximately 35 minutes. Students record their answers to questions on a machine-scored answer sheet that may be graded by hand or by machine. To each statement, students respond to one of five options: R--Rarely (0%-15%), S--Sometimes (16%-35%), F--Frequently (36%-65%), G--Generally (66%-85%), and A--Almost Always (86%-100%).

Four 25-item subtests are delineated. They are delay avoidance, work methods, teacher approval, and education acceptance. Scores on the first two subtests may be combined to yield a score for study habits. The last two subtests combine to form study attitudes. The total score is labeled "study orientation." Items under "delay avoidance," for example, ask students how much they stick to work until it is accomplished, how much their study is determined by mood, and how able they are to attend to their work. "Work-methods" items query about such habits and abilities as paper writing, test taking, and reading. "Teacher approval" presents statements to the student such as "Teachers enter teaching because they enjoy teaching" and "Teachers are sarcastic toward poorer students" (Brown & Holtzman, 1965, pp. 3-4). "Education acceptance" presents statements that deal with the reasons students are in college and the behaviors that students may exhibit as a result of their reasons for attending college.

Responses on the SSHA are based on student perception. The manual suggests that the administrator should encourage the students to express their true feelings about themselves (Brown & Holtzman, 1967). The self-rating technique on this survey has been listed as a threat to the instrument's predictive validity by Wrenn (1959) and

Shay (1972). Goldfried and D'Zurilla (1973), however, found a .55 correlation between the SSHA self-report and peer ratings. In establishing the concurrent validity of the instrument, Morris (1961) compared teacher ratings of academic performance and SSHA scores. Students who earned higher percentile ranks on the SSHA were rated as superior students by their teachers, and those students who were rated as poor students by their teachers had lower SSHA scores.

The manual also presents the evidence for the validity of the survey in terms of low (.21) mean correlation coefficients between the survey and aptitude tests and the somewhat higher mean correlation of .36 between the SSHA and grades. The survey measures something different from ability and appears to be somewhat more related to grades than to ability (Brown & Holtzman, 1967). Wen and Lui (1976) also found moderate, but significant, correlations (.34 to .49) between course examination grades and the SSHA.

Test-retest reliability coefficients taken on the SSHA at four-week intervals ranged from .93 to .95 for subtests and total scores. These data were collected using 237 ninth graders who were students in a Texas high school. More extensive reliability studies were conducted with Form H, which is the form for grades 7 through 12 (Brown & Holtzman, 1967).

Evaluators of the instrument generally concur that the instrument has been carefully devised and has satisfactory reliability and adequate validity (Deese, 1959; Higgins, 1967; Shay, 1972; Wrenn, 1959). However, the findings of this study are limited by the self-reporting format and the validity and reliability of the instrument.

The final examination on which the changed answers were observed was an 80-item comprehensive objective examination. It contained 40 multiple-choice items and 40 true-false items. Item 12 in the multiple-choice section was eliminated from the study because it was constructed with seven foils. The other multiple-choice items were constructed with four foils. This 80-item section was one of three parts of the examination.

The examination was analyzed at Michigan State University. The mean item difficulty was 29, and the mean item discrimination was 23. Of the 79 items on the examination, 45 items had a discrimination index of 23 or higher. The raw scores on the examination ranged from 36 to 74; the mean was 55.84 and the standard deviation was 7.52. The variance was 56.6. The Kuder-Richardson formula #20 reliability was .78.

Hypotheses

Ho 1: There is no effect due to directions on the number of answers changed.

Ho 2: There is no effect due to directions on the percentage of correct changes.

Ho 2.1: There is no effect due to directions on the percentage of correct changes on total items.

Ho 2.2: There is no effect due to directions on the percentage of correct changes on easy items.

Ho 2.3: There is no effect due to directions on the percentage of correct changes on difficult items.

Ho 3: There are no differences in SSHA scores among direction groups and effectiveness groups.

Ho 3.1: There are no differences in SSHA delay avoidance scores among direction groups and effectiveness groups on total items.

Ho 3.2: There are no differences in SSHA delay avoidance scores among direction groups and effectiveness groups on easy items.

Ho 3.3: There are no differences in SSHA delay avoidance scores among direction groups and effectiveness groups on difficult items.

Ho 3.4: There are no differences in SSHA work methods scores among direction groups and effectiveness groups on total items.

Ho 3.5: There are no differences in SSHA work methods scores among direction groups and effectiveness groups on easy items.

Ho 3.6: There are no differences in SSHA work methods scores among direction groups and effectiveness groups on difficult items.

Ho 3.7: There are no differences in SSHA education acceptance scores among direction groups and effectiveness groups on total items.

Ho 3.8: There are no differences in SSHA education acceptance scores among direction groups and effectiveness groups on easy items.

Ho 3.9: There are no differences in SSHA education acceptance scores among direction groups and effectiveness groups on difficult items.

Ho 3.10: There are no differences in SSHA teacher approval scores among direction groups and effectiveness groups on total items.

Ho 3.11: There are no differences in SSHA teacher approval scores among direction groups and effectiveness groups on easy items.

Ho 3.12: There are no differences in SSHA teacher approval scores among direction groups and effectiveness groups on difficult items.

Ho 3.13: There are no differences in SSHA composite score study habits (delay avoidance and work methods) among direction groups and effectiveness groups on total items.

Ho 3.14: There are no differences in SSHA composite score study habits (delay avoidance and work methods) among direction groups and effectiveness groups on easy items.

Ho 3.15: There are no differences in SSHA composite score study habits (delay avoidance and work methods) among direction groups and effectiveness groups on difficult items.

Ho 3.16: There are no differences in SSHA composite score study attitudes (education acceptance and teacher approval) among direction groups and effectiveness groups on total items.

Ho 3.17: There are no differences in SSHA composite score study attitudes (education acceptance and teacher approval) among direction groups and effectiveness groups on easy items.

Ho 3.18: There are no differences in SSHA composite score study attitudes (education acceptance and teacher approval) among direction groups and effectiveness groups on difficult items.

Ho 3.19: There are no differences in SSHA study orientation scores among direction groups and effectiveness groups on total items.

Ho 3.20: There are no differences in SSHA study orientation scores among direction groups and effectiveness groups on easy items.

Ho 3.21: There are no differences in SSHA study orientation scores among direction groups and effectiveness groups on difficult items.

Ho 4: There is no relationship between item difficulty and effective answer changing.

Ho 5: There are no differences in percentage of correct changes among direction groups and sex groups.

Ho 5.1: There are no differences in percentage of correct changes among direction groups and sex groups on total items.

Ho 5.2: There are no differences in percentage of correct changes among direction groups and sex groups on easy items.

Ho 5.3: There are no differences in percentage of correct changes among direction groups and sex groups on difficult items.

Ho 6: There are no differences in achievement measures among direction groups and effectiveness groups.

Ho 6.1: There are no differences in objective test scores among direction groups and effectiveness groups on total items.

Ho 6.2: There are no differences in objective test scores among direction groups and effectiveness groups on easy items.

Ho 6.3: There are no differences in objective test scores among direction groups and effectiveness groups on difficult items.

Ho 6.4: There are no differences in grade point averages among direction groups and effectiveness groups on total items.

Ho 6.5: There are no differences in grade point averages among direction groups and effectiveness groups on easy items.

Ho 6.6: There are no differences in grade point averages among direction groups and effectiveness groups on difficult items.

Ho 7: There are no differences in ability measures among direction groups and effectiveness groups.

Ho 7.1: There are no differences in SAT verbal scores among direction groups and effectiveness groups on total items.

Ho 7.2: There are no differences in SAT verbal scores among direction groups and effectiveness groups on easy items.

Ho 7.3: There are no differences in SAT verbal scores among direction groups and effectiveness groups on difficult items.

Ho 7.4: There are no differences in SAT mathematics scores among direction groups and effectiveness groups on total items.

Ho 7.5: There are no differences in SAT mathematics scores among direction groups and effectiveness groups on easy items.

Ho 7.6: There are no differences in SAT mathematics scores among direction groups and effectiveness groups on difficult items.

Data Analysis

The information obtained from answer changing, the SSHA scores, the item analysis, grade point averages, and SAT scores was entered into the computer at Michigan State University. The hypotheses were tested with the analysis of variance and the chi-square procedures, using the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1978). The results of the data analysis are discussed in Chapter IV.

Summary

This chapter contained a discussion of the methods and procedures used in the study. The population was described, the data-gathering procedures were outlined, and the design of the study was presented. In addition, the hypotheses were presented and the statistical treatment of the data was explained.

CHAPTER IV

PRESENTATION AND ANALYSIS OF DATA

Introduction

The purpose of this study was to investigate the effects of three different sets of directions to change answers on the answer changing behavior of college students. In this chapter, the statistical analysis of the data related to the major research questions is presented. The writer attempted to answer the following questions:

1. What is the effect of directions on the number of answers changed?

2. What is the effect of the directions on the percentage of correct changes?

3. Are there differences in SSHA scores among the direction groups (treatment) and the effectiveness groups (countereffective, ineffective, and effective changers)?

4. Is there a difference between easy and difficult item changes and the effectiveness of the change?

5. Are there any differences in the percentage of correct changes among the direction groups and the groups formed by sex of subjects?

6. Are there any differences in the achievement measures among the direction groups and the effectiveness groups?

7. Are there any differences in the ability measures among the direction groups and the effectiveness groups?

In Chapter III, the null hypotheses were presented. The hypotheses are restated in this chapter, and the findings are presented. The analysis of variance and the chi-square procedures were used to test the hypotheses.

Statistical Analysis

Directions and Number of Changes

Ho 1: There is no effect due to directions on the number of answers changed.

The mean numbers of answer changes were calculated for each group. These means are presented in Table 1.

Table 1.--Mean number of changes and standard deviations for each group.

Group	Size	Mean	Standard Deviation
Control	45	2.42	2.44
Treatment 1	40	2.20	1.77
Treatment 2	44	3.30	2.06

The analysis of variance failed to reject the hypothesis ($p > .05$). There were no differences in the mean number of changes due to the directions groups; $F(2, 123\ df) = 1.99$, $p > .05$. (See Table 2.) The directions given to the groups did not affect the number of changes.

Table 2.--ANOVA table for number of changes.

Source	df	MS	F
Treatment	2	9.189	1.99
Error	123	4.613	
Total	128	587.395	

$p > .05$

Directions and Effectiveness of Changes

Ho 2.1: There is no effect due to directions on the percentage of correct changes on total items.

The percentage of correct changes was used as a measure of answer changing effectiveness. This score was computed by dividing the number of correct changes by the total number of changes and then multiplying by 100. The range of percentage of correct changes, therefore, was 0-100 percent. This percentage-of-correct-changes score was calculated for total items, for easy items, and for difficult items. These divisions were made in the data because the test used as the data-gathering instrument was judged as easy (lower standard deviation than is suggested by testing authorities). The percentage correct score was used because the net-gain scores used in other studies on answer changing did not account for the total number of attempted changes. Net-gain changes are calculated by subtracting the incorrect changes from the correct changes. This procedure tends to equate individuals who have the same net-gain score when, in fact, one may have changed 18 items and the other, 4.

Easy items were those where 29 percent or fewer of the students missed the question. Difficult items were those where 30 percent or more of the students missed the question. The phrases "effective answer changing on total items" or "degree of change effectiveness on total items" refer to all correct items changed by the subjects.

To test the null hypothesis, the mean percentage correct for total items was calculated for each group. Table 3 lists these means.

Table 3.--Mean percentage correct scores and standard deviations on total items.

Group	Mean Percent Correct	Standard Deviation
Control	40.24	41.57
Treatment 1	50.87	56.23
Treatment 2	56.63	36.75

The analysis of variance failed to reject the null hypothesis ($p > .05$). There was no relationship between the means and the directions. No significant differences were found among these means; F (2, 123 df) = 1.60, $p > .05$. (See Table 4.) The directions did not affect the percentage of correct changes on total items.

Table 4.--ANOVA table for effectiveness of changes on total items.

Source	df	MS	F
Treatment	2	3219.16	1.60
Error	123	2007.56	
Total	128	2059.72	

$p > .05$.

Ho 2.2: There is no effect due to directions on the percentage of correct changes on easy items.

The mean percentage correct scores on easy items were calculated for students in each group. These means are presented in Table 5

Table 5.--Mean percentage correct scores and standard deviations on easy items.

Group	Mean Percent Correct	Standard Deviation
Control	37.34	43.57
Treatment 1	32.65	43.22
Treatment 2	61.74	44.72

The analysis of variance rejected the null hypothesis ($p < .05$); F (2, 120 df) = 5.39, $p < .05$. (See Table 6.) The directions did affect the percentage of correct changes on easy items. It would appear that the strong directions affected the percentage of correct changes. The mean percentage correct in treatment 2 was 29 percent higher than that in the treatment 1 group.

Table 6.--ANOVA table for effectiveness of changes on easy items.

Source	df	MS	F
Treatment	2	10483.300	5.39*
Error	120	1943.354	
Total	125	2059.254	

*$p < .05$.

Ho 2.3: There is no effect due to directions on the percentage of correct changes on difficult items.

The mean percentage correct for difficult-item changes was calculated for each group. These are listed in Table 7.

Table 7.--Mean percentage correct scores and standard deviations on difficult items.

Group	Mean Percent Correct	Standard Deviation
Control	26.88	42.25
Treatment 1	43.39	44.27
Treatment 2	35.03	39.80

The analysis of variance failed to reject the null hypothesis ($p > .05$); F (2, 123 df) = 1.60, $p > .05$. (See Table 8.) The directions did not affect the percentage of correct changes on difficult items. The mean percentage correct scores of students on difficult items were not significantly different among groups. The directions to change answers did not affect the percentage of correct changes on difficult items.

Table 8.--ANOVA table for effectiveness of changes on difficult items.

Source	df	MS	F
Treatment	2	2891.18	1.60
Error	123	1800.58	
Total	128	1788.37	

$p > .05$.

In summary, the directions to change answers did not affect the rate of answer changing or the percentage correct changes on total items and difficult items. Directions to change answers did statistically affect the percentage of correct changes on easy items.

The SSHA and Effectiveness of Changes

Ho 3.1: There are no differences in SSHA delay avoidance scores among direction groups and effectiveness groups on total items.

The SSHA delay avoidance subtest means were calculated by the directions groups and by the effectiveness groups. These means are presented in Table 9.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the effectiveness groups on total items; F (4, 119 df) = .47, $p > .05$. There was no main effect for directions; F (2, 119 df) = .398, $p > .05$. There was no main effect for effectiveness on total items; F (2, 119 df) = 1.57, $p > .05$. (See Table 10.) The directions and the degree of change effectiveness were not associated with the SSHA delay avoidance scores.

Table 9.--Mean and standard deviation (SD) scores for delay avoidance (DA) subtest for degree of change effectiveness on total items.

Group	DA--Total Countereffective			DA--Total Ineffective			DA--Total Effective			Total DA--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	18.52	9.03	5	27.40	7.16	17	20.70	6.39	45	20.33	8.23
Treatment 1	19	17.36	9.82	7	20.42	10.08	13	19.69	10.22	39	18.69	9.82
Treatment 2	11	18.72	9.49	11	20.18	11.86	22	20.90	7.33	44	20.18	9.36

Table 10.--ANOVA table for delay avoidance--total.

Source	df	MS	F
Treatment group	2	33.79	.39
Degree of effectiveness	2	133.70	1.57
Treatment x Degree	4	40.60	.47
Error	119	84.91	
Total	127	83.47	

$p > .05$

Ho 3.2: There are no differences in SSHA delay avoidance scores among direction groups and effectiveness groups on easy items.

The mean delay avoidance subtest scores were calculated by directions groups and by effectiveness groups on easy items. These means and standard deviations are presented in Table 11.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the effectiveness groups on easy items; F (4, 118 df) = .71, $p > .05$. There was no main effect for directions; F (2, 118 df) = .17, $p > .05$. There was no main effect for effectiveness; F (2, 118 df) = .31, $p > .05$. (See Table 12.) The delay avoidance scores were not associated with directions or with the degree of change effectiveness.

Table 11.--Mean and standard deviation (SD) scores for delay avoidance (DA) subtest for degree of change effectiveness on easy items.

Group	DA--Easy Countereffective			DA--Easy Ineffective			DA--Easy Effective			Total DA--Easy		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	21.44	8.72	5	16.75	10.43	17	19.21	6.64	45	20.33	8.23
Treatment 1	19	18.86	9.44	7	23.75	14.29	13	18.72	9.27	39	19.34	9.27
Treatment 2	11	21.53	11.02	11	15.00	1.13	22	20.00	9.00	44	20.18	9.49

Table 12.--ANOVA table for delay avoidance--easy.

Source	df	MS	F
Treatment group	2	14.09	.17
Degree of effectiveness	2	27.22	.32
Treatment x Degree	4	60.89	.71
Error	118	85.51	
Total	126	82.63	

$p > .05$

Ho 3.3: There are no differences in SSHA delay avoidance scores among direction groups and effectiveness groups on difficult items.

The mean delay avoidance scores were calculated by directions groups and by effectiveness groups on difficult items. These means and standard deviations are presented in Table 13. Also presented are the number of students in each degree-of-change-effectiveness category.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between directions groups and effectiveness groups; F (4, 120 df) = .37, $p > .05$. There was no main effect for directions; F (2, 120 df) = .48, $p > .05$. There was no main effect for effectiveness; F (2, 120 df) = 1.06, $p > .05$. (See Table 14.) Delay avoidance scores were not associated with the directions or with the effectiveness of changing.

Table 13.--Mean and standard deviation (SD) scores for delay avoidance (DA) subtest for degree of change effectiveness on difficult items.

Group	DA--Difficult Countereffective			DA--Difficult Ineffective			DA--Difficult Effective			Total DA--Difficult		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	19.25	8.83	3	26.00	8.18	11	21.81	5.90	45	20.33	8.23
Treatment 1	20	18.75	9.76	5	17.40	9.74	15	19.53	9.85	40	18.87	9.70
Treatment 2	26	18.75	9.15	7	22.71	10.91	11	22.00	10.13	44	20.18	9.49

Table 14.--ANOVA table for delay avoidance--difficult.

Source	df	MS	F
Treatment group	2	41.35	.48
Degree of effectiveness	2	91.17	1.06
Treatment x Degree	4	31.92	.37
Error	120	85.64	
Total	128	83.127	

$p > .05$.

In summary, for this population, when delay avoidance subtest scores were taken into account, there were no differences among direction groups or effectiveness groups on total items, easy items, or difficult items.

Ho 3.4: There are no differences in SSHA work methods scores among direction groups and effectiveness groups on total items.

The mean and standard deviation work methods scores were calculated by directions groups and by effectiveness groups on total items. These are presented in Table 15.

The analysis of variance failed to reject the hypothesis ($p > .05$). There was no interaction between the directions groups and the effectiveness groups; F (4, 119 df) = .16, $p > .05$. There was no main effect for directions; F (2, 119 df) = .50, $p > .05$. There was no main effect for effectiveness; F (2, 119 df) = 2.45, $p > .05$. (See Table 16.) Work methods scores were not associated with directions or with the degree of change effectiveness on total items.

Table 15.--Mean and standard deviation (SD) scores for work methods (WM) subtest for degree of change effectiveness on total items.

Group	WM--Total Countereffective			WM--Total Ineffective			WM--Total Effective			Total WM--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	24.08	10.37	5	28.80	6.26	17	27.00	7.65	45	25.71	9.04
Treatment 1	19	23.36	8.49	7	27.71	9.26	13	23.23	11.69	39	24.10	9.68
Treatment 2	11	23.18	8.30	11	29.00	7.38	22	23.95	11.95	44	25.02	9.22

Table 16.--ANOVA table for work methods--total.

Source	df	MS	F
Treatment group	2	44.05	.50
Degree of effectiveness	2	213.42	2.45
Treatment x Degree	4	14.39	.17
Error	119	86.91	
Total	127	85.68	

$p > .05$.

Ho 3.5: There are no differences in SSHA work methods scores among direction groups and effectiveness groups on easy items.

The mean and standard deviation work methods scores were calculated by directions groups and by effectiveness groups on easy items. These are presented in Table 17.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There was no interaction between the directions groups and the effectiveness groups on easy items; F (4, 118 df) = 2.06, $p > .05$. There was no main effect for directions; F (2, 118 df) = .30, $p > .05$. There was no main effect for effectiveness on easy items; F (2, 118 df) = 1.73, $p > .05$. Work methods scores were not associated with the directions or with the degree of change effectiveness on easy items.

Table 17.--Mean and standard deviation (SD) scores for work methods (WM) subtest for degree of change effectiveness on easy items.

Group	WM--Easy Countereffective			WM--Easy Ineffective			WM--Easy Effective			Total WM--Easy		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	27	26.07	8.42	4	24.75	9.14	14	25.28	8.89	45	25.71	9.04
Treatment 1	23	22.86	8.28	4	36.75	8.34	11	23.09	9.86	38	24.39	9.55
Treatment 2	15	28.46	11.35	3	24.33	8.38	26	23.11	7.59	44	25.02	9.22

Table 18.--ANOVA table for work methods--easy.

Source	df	MS	F
Treatment group	2	25.05	.31
Degree of effectiveness	2	145.20	1.77
Treatment x Degree	4	168.98	2.06
Error	118	81.95	
Total	126	84.70	

P > .05.

Ho 3.6: There are no differences in SSHA work methods scores among direction groups and effectiveness groups on difficult items.

The means and standard deviations were calculated for work methods subtest scores by directions groups and effectiveness groups on difficult items. (See Table 19.)

The analysis of variance failed to reject the hypothesis ($p > .05$). There was no interaction between the directions and the effectiveness; F (4, 120 df) = .86, $p > .05$. There was no main effect for directions; F (2, 120 df) = .65, $p > .05$. There was a main effect for effectiveness; F (2, 120 df) = 3.11, $p < .05$. (See Table 20.) The work methods scores were associated with effective answer changing on difficult items.

In summary, the SSHA work methods subtest was not associated with direction groups or effectiveness of change on easy items or total items. There was a relationship between the work methods subtest and the effectiveness of change on difficult items. Subjects

Table 19.--Mean and standard deviation (SD) scores for work methods (WM) subtest for degree of change effectiveness on difficult items.

Group	WM--Difficult Countereffective			WM--Difficult Ineffective			WM--Difficult Effective			Total WM--Difficult		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	24.35	9.58	3	27.00	2.00	11	29.18	8.01	45	25.71	9.04
Treatment 1	20	24.05	8.12	5	23.80	12.65	15	24.46	10.94	40	24.17	9.57
Treatment 2	26	21.96	9.33	7	28.28	7.56	11	30.18	7.26	44	25.02	9.22

who were higher scorers on the work methods subtest were the more effective answer changers on difficult items.

Table 20.--ANOVA table for work methods--difficult.

Source	df	MS	F
Treatment group	2	55.00	.65
Degree of effectiveness	2	250.01	3.11*
Treatment x Degree	4	71.93	.86
Error	120	83.56	
Total	128	85.04	

*$p < .05$.

Ho 3.7: There are no differences in SSHA education acceptance scores among direction groups and effectiveness groups on total items.

Mean and standard deviation scores for education acceptance subtests were computed by directions groups and by effectiveness groups. These scores are presented in Table 21.

The analysis of variance failed to reject the hypothesis ($p > .05$). There was no interaction between the directions groups and the effectiveness groups; F (4, 119 df) = .47, $p > .05$. There was no main effect for directions; F (2, 119 df) = 1.92, $p > .05$. There was a main effect for effectiveness; F (2, 119 df) = 4.54, $p > .05$. (See Table 22.) The education acceptance subtests were associated with effectiveness of changing on total items.

Table 21.--Mean and standard deviation (SD) scores for education acceptance (EA) subtest for degree of change effectiveness on total items.

Group	EA--Total Countereffective			EA--Total Ineffective			EA--Total Effective			Total EA--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	28.04	8.71	19	35.40	5.59	11	33.47	6.37	45	30.91	8.04
Treatment 1	5	29.15	10.06	7	31.71	7.27	11	31.61	5.17	39	30.43	8.14
Treatment 2	17	24.72	15.42	13	33.36	6.78	22	28.18	9.16	44	28.61	10.80

Table 22.--ANOVA table for education acceptance--total.

Source	df	MS	F
Treatment group	2	153.67	1.92
Degree of effectiveness	2	363.55	4.54*
Treatment x Degree	4	38.28	.47
Error	119	79.96	
Total	127	82.87	

*p = .01.

Ho 3.8: There are no differences in SSHA education acceptance scores among direction groups and effectiveness groups on easy items.

The mean and standard deviation SSHA education acceptance subtest scores were computed by directions groups and effectiveness groups on easy items. These data are presented in Table 23.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There was no interaction between the directions groups and the effectiveness groups on easy items; F (4, 118 df) = .64; $p > .05$. There was no main effect for directions; F (2, 118 df) = .24, $p > .05$. There was no main effect for effectiveness; F (2, 118 df) = .14, $p > .05$. (See Table 24.) The mean education acceptance scores were not associated with directions or with effectiveness of change on easy items.

Table 23.--Mean and standard deviation (SD) scores for education acceptance (EA) subtest for degree of change effectiveness on easy items.

Group	EA--Easy Countereffective			EA--Easy Ineffective			EA--Easy Effective			Total EA--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	27	30.33	8.72	23	28.75	4.34	15	32.64	7.56	45	30.91	8.04
Treatment 1	4	28.69	9.08	4	34.50	4.04	3	33.54	4.96	38	30.71	7.97
Treatment 2	14	29.13	12.90	11	29.66	10.69	26	28.19	9.90	44	28.61	10.80

Table 24.--ANOVA table for education acceptance--easy.

Source	df	MS	F
Treatment group	2	96.17	1.14
Degree of effectiveness	2	49.30	.58
Treatment x Degree	4	55.57	.62
Error	118	83.92	
Total	126	82.25	

P > .05.

Ho 3.9: There are no differences in SSHA education acceptance scores among direction groups and effectiveness groups on difficult items.

The education acceptance mean and standard deviation scores were computed by directions groups and by effectiveness groups on difficult items. These data are presented in Table 25.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There was no interaction between directions and effectiveness on difficult items; F (4, 120 df) = .22, $p > .05$. There was no main effect for directions; F (2, 120 df) = .88, $p > .05$. There was no main effect for effectiveness; F (2, 120 df) = 2.69, $p > .05$. (See Table 26.)

In summary, the education acceptance subtest scores were associated with the effectiveness of change on total items. Education acceptance scores were not associated with effectiveness of change on easy items or difficult items.

Table 25.--Mean and standard deviation (SD) scores for education acceptance (EA) subtest for degree of change effectiveness on difficult items.

Group	EA--Difficult Countereffective			EA--Difficult Ineffective			EA--Difficult Effective			Total EA--Difficult		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	29.51	8.80	20	32.00	3.60	26	34.54	5.20	45	30.90	8.04
Treatment 1	3	28.85	8.42	5	33.60	7.80	7	31.66	7.62	40	30.50	8.05
Treatment 2	11	27.19	11.58	15	28.42	10.51	11	32.09	9.06	44	28.60	10.80

Table 26.--ANOVA table for education acceptance--difficult.

Source	df	MS	F
Treatment group	2	72.86	.88
Degree of effectiveness	2	222.33	2.69*
Treatment x Degree	4	18.30	.22
Error	120	82.36	
Total	128	82.29	

*p = .01.

Ho 3.10: There are no differences in SSHA teacher approval scores among direction groups and effectiveness groups on total items.

The means and standard deviations for teacher approval scores were calculated by directions groups and by effectiveness groups on total items. (See Table 27.) In all three treatment groups, the highest mean teacher approval scores were earned by the ineffective changers.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There was no interaction between directions groups and the effectiveness groups on total items; F (4, 119 df) = 1.08, $p > .05$. There were no main effects for directions; F (2, 119 df) = .35, $p > .05$. There was a main effect for effectiveness; F (2, 119 df) = 3.54, $p < .05$. (See Table 28.) The teacher approval means were associated with the degree of effectiveness of changes on total items. The subjects with higher mean teacher approval scores were the ineffective changers.

Table 27.--Mean and standard deviation (SD) scores for teacher approval (TA) subtest for degree of change effectiveness on total items.

Group	TA--Total Countereffective			TA--Total Ineffective			TA--Total Effective			Total TA--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	24.52	7.80	5	31.40	3.50	17	28.64	7.24	45	26.84	7.55
Treatment 1	19	27.84	8.38	7	33.42	5.62	13	24.61	8.51	39	27.76	8.39
Treatment 2	11	24.00	13.40	11	29.18	5.54	22	27.18	8.56	44	26.88	9.39

Table 28.--ANOVA table for teacher approval--total.

Source	df	MS	F
Treatment group	2	24.69	.35
Degree of effectiveness	2	244.01	3.54*
Treatment x Degree	4	75.04	1.08
Error	119	68.916	
Total	127	70.95	

*p = .032.

Ho 3.11: There are no differences in SSHA teacher approval scores among direction groups and effectiveness groups on easy items.

The mean teacher approval subtest scores and standard deviations were computed by directions groups and by effectiveness groups on easy items. These data are presented in Table 29.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between directions and effectiveness; F (4, 118 df) = .64, $p > .05$. There was no main effect for directions; F (2, 118 df) = .24, $p > .05$. There was no main effect for effectiveness; F (2, 118 df) = .14, $p > .05$. The teacher approval scores were not associated with directions or with change effectiveness on easy items.

Table 29.--Mean and standard deviation (SD) scores for teacher approval (TA) subtest for degree of change effectiveness on easy items.

Group	TA--Easy Countereffective			TA--Easy Ineffective			TA--Easy Effective			Total TA--Easy		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	27	26.85	7.87	4	25.00	3.36	14	27.35	8.08	25	26.84	7.55
Treatment 1	23	27.30	7.35	4	34.25	6.65	11	27.18	9.73	38	28.00	8.12
Treatment 2	15	26.60	11.87	3	25.00	6.55	26	27.26	8.29	44	26.88	9.39

Table 30.--ANOVA table for teacher approval--easy.

Source	df	MS	F
Treatment group	2	18.01	.24
Degree of effectiveness	2	10.40	.14
Treatment x Degree	4	46.73	.64
Error	118	72.42	
Total	126	69.75	

$p > .05$.

Ho 3.12: There are no differences in SSHA teacher approval scores among direction groups and effectiveness groups on difficult items.

The means and standard deviations for teacher approval scores were calculated by directions groups and by effectiveness groups. These data are presented in Table 31.

The analysis of variance failed to reject the ($p > .05$). There were no interactions between the directions groups and the effectiveness groups on difficult items; F (4, 120 df) = 2.63, $p > .05$. There was no main effect for directions; F (2, 120 df) = .07, $p > .05$. There was no main effect for effectiveness; F (2, 120 df) = 2.63, $p > .05$. (See Table 32.) The teacher approval scores were not associated with the directions or the change effectiveness on difficult items.

In summary, SSHA teacher approval subtest scores were associated with the effectiveness of change on total items. The teacher

Table 31.--Mean and standard deviation (SD) scores for teacher approval (TA) subtest for degree of change effectiveness on difficult items.

Group	TA--Difficult Countereffective			TA--Difficult Ineffective			TA--Difficult Effective			Total TA--Difficult		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	25.54	7.66	3	32.00	4.58	11	29.09	7.21	45	26.84	7.55
Treatment 1	20	26.80	6.90	5	33.60	6.87	15	27.00	9.55	40	27.72	8.29
Treatment 2	26	25.42	9.88	7	28.14	10.13	11	29.54	7.68	44	26.88	9.39

approval subtest was not associated with the degree of change effectiveness on easy or difficult items, and there were no effects for directions.

Table 32.--ANOVA table for teacher approval--difficult.

Source	df	MS	F
Treatment group	2	5.18	.07
Degree of effectiveness	2	186.08	2.63
Treatment x Degree	4	39.30	.56
Error	120	70.52	
Total	128	70.41	

$p > .05$.

Ho 3.13: There are no differences in SSHA composite score study habits (delay avoidance and work methods) among direction groups and effectiveness groups on total items.

The means and standard deviations for the study habits scores by directions groups and effectiveness groups on total items were calculated. These data are presented in Table 33.

The analysis of variance failed to reject the hypothesis ($p > .05$). There were no interactions between directions and effectiveness; F (4, 119 df) = .10, $p > .05$. There was no main effect for directions; F (2, 119 df) = .59, $p > .05$. There was no main effect for effectiveness of change; F (2, 119 df) = 2.29, $p > .05$. The SSHA means were not associated with the treatment or with the degree of change effectiveness on total items.

Table 33.--Mean and standard deviation (SD) scores for study habits (SH) subtest for degree of change effectiveness on total items.

Group	SH--Total Countereffective			SH--Total Ineffective			SH--Total Effective			Total SH--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	42.60	17.28	5	56.20	10.13	17	47.64	12.26	45	46.02	15.25
Treatment 1	19	40.21	17.00	7	48.00	18.95	13	42.98	21.43	39	42.51	18.63
Treatment 2	11	42.09	22.77	11	49.18	12.68	22	44.86	15.24	44	45.25	16.67

Table 34.--ANOVA table for study habits--total.

Source	df	MS	F
Treatment group	2	169.18	.59
Degree of effectiveness	2	653.63	2.29
Treatment x Degree	4	30.08	.11
Error	119	285.35	
Total	127	250.81	

$p > .05$.

Ho 3.14: There are no differences in SSHA composite score study habits (delay avoidance and work methods) among direction groups and effectiveness groups on easy items.

Study habits means and standard deviations were computed by directions groups and by effectiveness groups on easy items. These means and standard deviations are given in Table 35.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the effectiveness groups; F (4, 118 df) = 1.21, $p > .05$. There was no main effect for directions; F (2, 118 df) = .35, $p > .05$. There was no main effect for effectiveness; F (2, 118 df) = .60, $p > .05$. (See Table 35.) The directions and the degree of effectiveness of change were not associated with the study habits scores on easy item changes.

Table 35.--Mean and standard deviation (SD) scores for study habits (SH) subtest for degree of change effectiveness on easy items.

Group	SH--Easy Countereffective			SH--Easy Ineffective			SH--Easy Effective			Total SH--Easy		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	27	47.51	16.39	4	41.50	9.03	14	44.42	14.74	45	46.02	15.25
Treatment 1	23	41.73	16.83	4	58.00	25.62	11	41.72	18.62	38	43.44	18.46
Treatment 2	15	50.13	20.96	3	39.33	7.76	26	43.11	14.28	44	45.25	16.67

Table 36.--ANOVA table for study habits--easy.

Source	df	MS	F
Treatment group	2	98.07	.35
Degree of effectiveness	2	169.92	.60
Treatment x Degree	4	341.48	1.21
Error	118	280.54	
Total	126	277.397	

$p > .05$.

Ho 3.15: There are no differences in SSHA composite score study habits (delay avoidance and work methods) among direction groups and effectiveness groups on difficult items.

The means and standard deviations for the study habits composite score by directions groups and by effectiveness groups on difficult items were computed. These means are presented in Table 37.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There was no interaction between the directions groups and the effectiveness groups; F (4, 120 df) = .50, $p > .05$. There was no main effect for directions; F (2, 120 df) = .77, $p > .05$. There was no main effect for effectiveness; F (2, 120 df) = 2.40, $p > .05$. (See Table 38.) SSHA study habits scores were not associated with the directions or with the degree of change effectiveness on difficult items.

In summary, when the study habits scores were taken into account, there were no differences among direction groups or effectiveness groups on easy, difficult, or total items.

Table 37.--Mean and standard deviation (SD) scores for study habits (SH) subtest for degree of change effectiveness on difficult items.

Group	SH--Difficult Countereffective			SH--Difficult Ineffective			SH--Difficult Effective			Total SH--Difficult		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	43.61	16.19	3	53.00	10.14	11	50.90	12.54	45	46.02	15.25
Treatment 1	20	42.25	16.54	5	41.20	23.87	15	44.00	20.31	40	42.72	18.46
Treatment 2	26	40.76	16.37	7	51.00	16.91	11	52.00	15.03	44	45.25	16.67

Table 38.--ANOVA table for study habits--difficult.

Source	df	MS	F
Treatment group	2	217.16	.77
Degree of effectiveness	2	672.71	2.40
Treatment x Degree	4	141.42	.50
Error	120	279.84	
Total	128	279.15	

$p > .05$.

Ho 3.16: There are no differences in SSHA composite score study attitudes (education acceptance and teacher approval) among direction groups and effectiveness groups on total items.

The mean and standard deviation scores were computed for the study attitudes composite scores by directions groups and by effectiveness groups on total items. These data are presented in Table 39.

The analysis of variance failed to reject the null hypothesis. ($p > .05$). There were no interactions between the directions and effectiveness groups; F (4, 119 df) = .62, $p > .05$. There was no main effect for the directions; F (2, 119 df) = 1.30, $p > .05$. There was a main effect for effectiveness; F (2, 119 df) = 3.87, $p > .05$. (See Table 40.) There were significant differences among the study attitudes means. The study attitudes means were associated with the degree of change effectiveness on total items.

Table 39.--Mean and standard deviation (SD) scores for study attitudes (SA) subtest for degree of change effectiveness on total items.

Group	SA--Total Countereffective			SA--Total Ineffective			SA--Total Effective			Total SA--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	52.86	16.86	5	66.80	6.45	17	62.11	12.68	45	57.91	15.16
Treatment 1	19	56.89	17.86	7	61.85	14.66	13	55.38	13.69	39	57.28	15.79
Treatment 2	11	47.81	26.86	11	62.54	9.89	22	54.04	15.26	44	54.61	18.23

Table 40.--ANOVA table for study attitudes--total.

Source	df	MS	F
Treatment group	2	341.61	1.30
Degree of effectiveness	2	1016.74	3.87*
Treatment x Degree	4	162.56	.62
Error	119	262.28	
Total	127	269.01	

*$p < .05$.

Ho 3.17: There are no differences in SSHA composite score study attitudes (education acceptance and teacher approval) among direction groups and effectiveness groups on each items.

The means and standard deviations were computed for the study attitudes composite scores by directions groups and by effectiveness groups on easy items. These scores are reported in Table 41.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the effectiveness groups; F (4, 119 df) = .60, $p > .05$. There was no main effect by directions; F (2, 119 df) = .67, $p > .05$. There was no main effect by effectiveness of change on easy items; F (2, 119 df) = .35, $p > .05$. (See Table 42.) The study attitudes scores were not associated with directions or with change effectiveness on easy items.

Table 41.--Mean and standard deviation (SD) scores for study attitudes (SA) subtest for degree of change effectiveness on easy items.

Group	SA--Easy Countereffective			SA--Easy Ineffective			SA--Easy Effective			Total SA--Easy		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	27	57.44	16.37	4	53.75	6.94	14	60.00	14.83	45	57.91	15.16
Treatment 1	23	54.91	15.47	4	68.75	10.30	11	59.72	15.47	38	57.76	15.32
Treatment 2	15	55.06	23.08	3	54.66	16.25	26	54.34	15.88	44	54.61	18.23

Table 42.--ANOVA table for study attitudes--easy.

Source	df	MS	F
Treatment group	2	185.03	.67
Degree of effectiveness	2	95.73	.35
Treatment x Degree	4	165.45	.60
Error	118	273.31	
Total	126	265.12	

$p > .05$.

Ho 3.18: There are no differences in SSHA composite score study attitudes (education acceptance and teacher approval) among direction groups and effectiveness groups on difficult items.

The study attitudes mean and standard deviation scores were calculated by directions groups and by effectiveness groups on difficult items. (See Table 43.)

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between directions groups and effectiveness groups; F (4, 120 df) = .15, $p > .05$. There was no main effect by directions; F (2, 120 df) = .62, $p > .05$. There was no main effect by effectiveness of change on difficult items; F (2, 120 df) = 2.19, $p > .05$. (See Table 44.) Study attitudes scores were not associated with directions or with the degree of change effectiveness on difficult items.

In summary, the SSHA study attitudes composite scores were associated with the degree of change effectiveness on total items. There

Table 43.--Mean and standard deviation (SD) scores for study attitudes (SA) subtest for degree of change effectiveness on difficult items.

Group	SA--Difficult Countereffective			SA--Difficult Ineffective			SA--Difficult Effective			Total SA--Difficult		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	55.29	16.47	3	64.00	7.21	11	63.63	10.89	45	57.91	15.16
Treatment 1	20	55.00	16.92	5	62.80	16.52	15	58.60	16.65	40	57.32	15.59
Treatment 2	26	52.23	19.84	7	56.57	18.55	11	59.00	14.13	44	54.61	18.23

was no association between the composite scores and the degree of change effectiveness for easy items and difficult items.

Table 44.--ANOVA table for study attitudes--difficult.

Source	df	MS	F
Treatment group	2	170.60	.63
Degree of effectiveness	2	595.30	2.19
Treatment x Degree	4	40.64	.15
Error	120	271.21	
Total	128	266.96	

$p > .05$.

Ho 3.19: There are no differences in SSHA study orientation scores among direction groups and effectiveness groups on total items.

The mean and standard deviation study orientation scores were computed by directions groups and by effectiveness groups on total items. These means are presented in Table 45. In all three groups, the means for ineffective changers were the highest means.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the effectiveness groups; F (4, 119 df) = .25, $p > .05$. There was no main effect for directions; F (2, 119 df) = .68, $p > .05$. There was a main effect for effectiveness of change; F (2, 119 df) = 3.34, $p < .05$. (See Table 46.) The study orientation scores were associated with

Table 45.--Mean and standard deviation (SD) scores for study orientation (SO) subtest for degree of change effectiveness on total items.

Group	SO--Total Countereffective			SO--Total Ineffective			SO--Total Effective			Total DO--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	95.52	32.47	5	123.00	15.44	17	109.76	23.31	45	103.95	28.95
Treatment 1	19	97.15	37.11	7	110.00	31.23	13	98.30	32.32	39	99.84	31.55
Treatment 2	11	90.18	49.68	11	111.72	19.93	22	98.90	27.28	44	100.09	32.97

the degree of change effectiveness on total items. The more effective changers had higher mean study orientation scores.

Table 46.--ANOVA table for study orientation--total.

Source	df	MS	F
Treatment group	2	652.29	.68
Degree of effectiveness	2	3209.42	3.34*
Treatment x Degree	4	240.01	.25
Error	119	958.83	
Total	127	960.18	

*p = .039.

Ho 3.20: There are no differences in SSHA study orientation scores among direction groups and effectiveness groups on easy items.

The means and standard deviations were computed for the study orientation scores by directions groups and by effectiveness groups on easy items. These scores are presented in Table 47.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between directions and effectiveness of change; F (4, 118 df) = 1.00, $p > .05$. There was no main effect for directions; F (2, 118 df) = .14, $p > .05$. There was no main effect for effectiveness of change; F (2, 118 df) = .15, $p > .05$. (See Table 48.) The study orientation scores were not associated with directions or with the degree of change effectiveness, nor was there an interaction between these variables.

Table 47.--Mean and standard deviation (SD) scores for study orientation (SO) subtest for degree of change effectiveness on easy items.

Group	SO--Easy Countereffective			SO--Easy Ineffective			SO--Easy Effective			Total SO--Easy		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	27	105.00	31.41	4	95.25	15.71	14	104.42	27.93	45	103.95	28.94
Treatment 1	23	96.73	28.89	4	127.75	35.30	11	101.45	31.54	38	101.20	30.81
Treatment 2	15	105.86	43.66	3	94.00	24.02	26	97.46	26.92	44	100.09	32.97

Table 48.--ANOVA table for study orientation--easy.

Source	df	MS	F
Treatment group	2	138.23	.14
Degree of effectiveness	2	148.37	.15
Treatment x Degree	4	974.92	1.00
Error	118	970.88	
Total	126	945.313	

$p > .05$.

Ho 3.21: There are no differences in SSHA study orientation scores among direction groups and effectiveness groups on difficult items.

The mean and standard deviation study orientation scores by directions groups and by effectiveness groups on difficult items are presented in Table 49.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between directions and effectiveness groups; F (4, 120 df) = .21, $p > .05$. There was no main effect for directions; F (2, 120 df) = .48, $p > .05$. There was no main effect for effectiveness; F (2, 120 df) = 2.56, $p > .05$. (See Table 50.)

Study orientation total scores were associated with the degree of change effectiveness on total items. The study orientation scores were not associated with the degree of change effectiveness on easy or on difficult items or with the directions.

Table 49.--Mean and standard deviation (SD) scores for study orientation (SO) subtest for degree of change effectiveness on difficult items.

Group	SO--Difficult Countereffective			SO--Difficult Ineffective			SO--Difficult Effective			Total SO--Difficult		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	98.93	31.25	3	117.00	17.00	11	114.00	21.12	45	103.90	28.90
Treatment 1	20	97.30	29.49	5	104.00	35.24	15	102.66	33.67	40	100.15	31.21
Treatment 2	26	93.38	34.78	7	107.57	34.01	11	111.18	25.80	44	100.09	32.97

Table 50.--ANOVA table for study orientation--difficult.

Source	df	MS	F
Treatment group	2	467.53	.48
Degree of effectiveness	2	2472.83	2.56
Treatment x Degree	4	206.33	.21
Error	120	965.44	
Total	128	953.56	

$p > .05$.

In summary, the education acceptance and teacher approval subtests and the composite score of those two subtests, study attitudes, were associated with the degree of change effectiveness on total items. The SSHA study orientation total score was also associated with change effectiveness on total items. The SSHA scores were not associated with change effectiveness on easy items. The degree of change effectiveness on difficult items was associated with the work methods subtest. It should be noted that the education acceptance, teacher approval, and study attitudes SSHA subtest scores and the total score, study orientation, were associated with the degree of change effectiveness on difficult items at the .10 level.

Item Difficulty and
Change Effectiveness

Ho 4: There is no relationship between item difficulty and effective answer changing.

A comparison was made between the number of students who were effective, ineffective, and countereffective changers on easy items and the number of students who were effective, ineffective, and countereffective changers on difficult items. These data are presented in Table 51.

Table 51.--Degree of change effectiveness on easy and difficult items.

Easy Items	Difficult Items			Total
	Countereffective	Ineffective	Effective	
Countereffective	42	8	18	68 52.7%
Ineffective	7	2	2	11 8.5%
Effective	28	5	17	50 38.8%
Total	77 59.7%	15 11.6%	37 28.7%	129 100.0%

$p < .05$.

The chi-square test failed to reject the null hypothesis ($p < .05$). There was no significant relationship between the effectiveness of changers on easy items and on difficult items X^2 (4 df) = 1.74, $p > .05$. Easy and difficult items were changed with the same degree of change effectiveness.

Sex Differences and Effective Changing

Ho 5.1: There are no differences in percentage of correct changes among direction groups and sex groups.

Effective answer changing was defined as the percentage of correct changes. The mean percentage of correct changes for total items for males and females in each direction group was calculated. These are presented in Table 52. Treatment 1 group females had a much higher mean percentage of correct changes on total items than males.

Table 52.--Percentage of correct changes on total items for males and females.

Group	Male Mean Percentage of Correct Changes			Female Mean Percentage of Correct Changes			Total Mean Percentage of Correct Changes		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	17	40.39	44.95	28	40.15	40.56	45	40.24	41.57
Treatment 1	16	31.25	34.35	24	63.95	64.38	40	50.87	56.23
Treatment 2	18	53.57	36.28	26	58.76	37.64	44	56.63	36.75

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the effectiveness groups on total items; F (2, 123 df) = 1.54, $p > .05$. There was no main effect for directions; F (1, 123 df) = 1.60, $p > .05$. There was no main effect for effectiveness of change; F (1, 123 df) = 2.17, $p > .05$. (See Table 53.) There were no differences in the percentage of correct changes due to sex or directions.

Table 53.--ANOVA table for percentage of correct changes on total items for males and females.

Source	df	MS	F
Treatment group	2	3219.16	1.60
Sex	1	4361.93	2.17
Treatment x Sex	2	3097.93	1.54
Error	123	2007.56	
Total	128	2059.72	

$p > .05$.

Ho 5.2: There are no differences in percentage of correct changes among direction groups and sex groups on easy items.

The mean percentage of correct changes on easy items was computed for males and females in each of the treatment groups. These data are presented in Table 54. Treatment 1 males had a higher mean percentage of correct changes on the easy items than did females. The control group and treatment 2 group mean percentages of correct changes for males and females on easy items appeared to be similar.

The analysis of variance failed to reject the hypothesis ($p > .05$). There were no interactions between the directions groups and the sex groups; F (2, 123 df) = .64, $p > .05$. There was no main effect for sex; F (1, 123 df) = .57, $p > .05$. There was a main effect for directions; F (1, 123 df) = 5.39, $p < .05$. (See Table 55.) The percentage of correct changes was affected by the directions. Easy item changes were affected by the strong directions to go back and revise answers.

Table 54.--Mean percentage of correct changes on easy items for males and females.

Group	Male Mean Percentage of Correct Changes			Female Mean Percentage of Correct Changes			Total Mean Percentage of Correct Changes		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	17	36.60	44.18	28	37.79	44.00	45	37.34	43.57
Treatment 1	13	39.93	32.52	24	19.23	47.07	37	32.65	43.22
Treatment 2	18	62.03	43.83	26	61.53	46.19	44	61.74	44.72

Table 55.--ANOVA table for percentage of correct changes on easy items for males and females.

Source	df	MS	F
Treatment group	2	10483.30	5.39*
Sex	1	1116.076	.57
Treatment x Sex	2	1257.41	.64
Error	120	1943.35	
Total	125	2059.25	

*$p < .05$.

Ho 5.3: There are no differences in percentage of correct changes among direction groups and sex groups on difficult items.

Mean percentage of correct changes on difficult items was calculated for males and for females by directions groups. These data are presented in Table 56.

Table 56.--Mean percentage of correct changes and standard deviations on difficult items for males and females by treatment group.

Group	Male Mean Percentage of Correct Changes			Female Mean Percentage of Correct Changes			Total Mean Percentage of Correct Changes		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	17	20.58	39.76	28	30.71	43.96	45	26.88	42.25
Treatment 1	16	47.91	46.69	24	40.38	43.33	40	43.39	44.27
Treatment 2	18	36.11	40.82	26	34.29	39.88	44	35.03	39.80

The analysis of variance failed to reject the hypothesis ($p > .05$). There were no significant interactions between the directions groups and the sex groups; F (2, 123 df) = .46, $p > .05$. There was no main effect for directions; F (1, 123 df) = 1.60, $p > .05$. There was no main effect for sex; F (1, 123 df) = .004, $p > .05$. (See Table 57.) The directions and sex of subjects did not affect and were not associated with the percentage of correct changes on difficult items.

Table 57.--ANOVA table for percentage of correct changes on difficult items for males and females.

Source	df	MS	F
Treatment group	2	2891.13	1.60
Sex	1	7.80	.004
Treatment x Sex	2	828.09	.46
Error	123	1800.58	
Total	128	1788.37	

$p > .05$.

In summary, the degree of percentage of correct changes on total items, easy items, or difficult items was not associated with the sex of the subjects, nor was it affected by the directions to change. Although there were some differences within treatment groups, these differences were not significant. The mean percentages of correct changes for males and females did not differ significantly.

Achievement and Answer Changing

Ho 6.1: There are no differences in the objective test scores among direction groups and effectiveness groups on total items.

The means and standard deviations for the final examination by directions groups and by effectiveness groups on total items were calculated. These means are presented in Table 58.

The analysis of variance failed to reject the null hypothesis. ($p > .05$). There were no interactions between the directions groups and the effectiveness groups; F (4, 119 df) = .28, $p > .05$. There was no main effect for directions; F (2, 119 df) = 1.07, $p > .05$. There was no main effect for effectiveness; F (2, 119 df) = .84, $p > .05$. (See Table 59.) Objective test scores were not associated with directions or with the degree of change effectiveness on total items.

Table 58.--Means and standard deviations for objective test scores for the degree of change effectiveness on total items.

Group	Mean Test Score--Total Countereffective			Mean Test Score--Total Ineffective			Mean Test Score--Total Effective			Total Mean Test Score--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	55.26	6.28	5	60.40	3.64	17	57.47	6.72	45	56.66	6.34
Treatment 1	19	55.21	8.82	7	57.14	4.09	13	57.07	9.88	39	56.17	8.44
Treatment 2	11	54.90	9.65	11	55.18	8.79	22	54.59	6.36	44	54.81	7.71

Table 59.--ANOVA table for objective final--total.

Source	df	MS	F
Treatment group	2	61.91	1.07
Degree of effectiveness	2	49.02	.85
Treatment x Degree	4	16.39	.28
Error	119	57.80	
Total	127	56.089	

$p > .05$.

Ho 6.2: There are no differences in objective test scores among direction groups and effectiveness groups on easy items.

The mean and standard deviation scores for the effectiveness groups and the directions groups on easy items are presented in Table 60.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the effectiveness groups; F (4, 118 df) = .52, $p > .05$. There was no main effect for directions; F (2, 118 df) = .66, $p > .05$. There was no main effect for effectiveness; F (2, 118 df) = .09, $p > .05$. (See Table 61.) Objective test scores were not associated with directions or with change effectiveness on easy items.

Table 60.--Means and standard deviations for objective test scores for the degree of change effectiveness on easy items.

Group	Mean Test Score--Easy Countereffective			Mean Test Score--Easy Ineffective			Mean Test Score--Easy Effective			Total Mean Test Score--Easy		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	27	56.74	6.26	4	54.25	7.58	14	57.21	6.51	45	56.66	6.34
Treatment 1	23	54.95	7.96	4	57.25	2.98	11	57.63	11.67	38	55.97	8.76
Treatment 2	15	56.20	10.03	3	53.00	3.60	26	54.23	6.56	44	54.81	7.77

Table 61.--ANOVA table for objective final score--easy.

Source	df	MS	F
Treatment group	2	39.85	.67
Degree of effectiveness	2	5.81	.10
Treatment x Degree	4	31.085	.52
Error	118	59.67	
Total	126	57.57	

$p > .05$.

Ho 6.3: There are no differences in objective test scores among direction groups and effectiveness groups on difficult items.

The mean and standard deviation objective test scores were calculated for directions groups and for effectiveness groups on difficult items. These means are presented in Table 62.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the effectiveness groups; F (4, 120 df) = .806, $p > .05$. There was no main effect for directions; F (2, 120 df) = .97, $p > .05$. There was a main effect for effectiveness of change; F (2, 120 df) = 4.78, $p < .05$. Objective test scores were associated with the degree of effectiveness of changes on difficult items. (See Table 63.) There was a difference among the mean objective test scores for counter-effective, ineffective, and effective changers.

Table 62.--Means and standard deviations for objective test scores for the degree of change effectiveness on difficult items.

Group	Mean Test Score--Difficult Countereffective			Mean Test Score--Difficult Ineffective			Mean Test Score--Difficult Effective			Total Mean Test Score--Difficult		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	55.58	6.05	3	59.33	3.21	11	59.00	6.34	45	56.66	6.34
Treatment 1	20	52.85	7.45	5	60.20	7.94	15	58.46	9.13	40	55.87	8.55
Treatment 2	26	53.76	7.76	7	53.00	9.81	11	58.45	5.26	44	54.81	7.71

Table 63.--ANOVA table for objective final--difficult.

Source	df	MS	F
Treatment group	2	52.82	.97
Degree of effectiveness	2	258.86	4.78*
Treatment x Degree	4	43.63	.80
Error	120	54.12	
Total	128	56.74	

*p = .01.

Ho 6.4: There are no differences in grade point averages among direction groups and effectiveness groups on total items.

Grade point averages by directions groups and by effectiveness groups were computed on total items. These are presented in Table 64.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions and the effectiveness of changes on total items; F (4, 119 df) = 1.62, $p > .05$. There was no main effect for directions; F (2, 119 df) = .77, $p > .05$. There was no main effect for effectiveness; F (2, 119 df) = .19, $p > .05$. The grade point averages were not associated with directions or with effectiveness of changes on total items.

Table 64.--Mean student grade point averages (GPA) and standard deviations (SD) for degree of change effectiveness on total items.

Group	Mean GPA-- Total Countereffective			Mean GPA-- Total Ineffective			Mean GPA-- Total Effective			Total Mean GPA-- Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	2.47	.60	5	3.25	.35	17	2.68	.74	45	2.64	.74
Treatment 1	19	2.56	.88	7	2.39	.82	13	2.84	1.00	39	2.45	.89
Treatment 2	11	2.44	.71	11	2.25	.63	22	2.63	.56	44	2.48	.62

Table 65.--ANOVA table for grade point average--total.

Source	df	MS	F
Treatment group	2	44.12	.77
Degree of effectiveness	2	11.15	.19
Treatment x Degree	4	94.09	1.65
Error	119	56.95	
Total	127	57.16	

$p > .05$.

Ho 6.5: There are no differences in grade point averages among direction groups and effectiveness groups on easy items.

The mean grade point averages by direction groups and by effectiveness groups on easy items were calculated. (See Table 66.)

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions and the change effectiveness on easy items; F (4, 118 df) = 1.22, $p > .05$. There was no main effect for directions; F (2, 118 df) = .99, $p > .05$. There was no main effect for change effectiveness; F (2, 118 df) = 1.17, $p > .05$. (See Table 67.) Grade point averages were not associated with directions or change effectiveness on easy items.

Table 66.--Mean grade point averages (GPA) and standard deviations (SD) for degree of change effectiveness on easy items.

Group	Mean GPA-- Easy Countereffective			Mean GPA-- Easy Ineffective			Mean GPA-- Easy Effective			Total Mean GPA-- Easy		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	27	2.68	.64	4	2.18	.77	14	2.86	.90	45	2.64	.74
Treatment 1	23	2.29	.89	4	2.64	.95	11	2.63	.88	38	2.42	.89
Treatment 2	15	2.59	.62	3	1.69	.83	26	2.52	.56	44	2.48	.62

Table 67.--ANOVA table for grade point average--easy.

Source	df	MS	F
Treatment group	2	55.69	.99
Degree of effectiveness	2	65.79	1.17
Treatment x Degree	4	68.79	1.22
Error	118		
Total	126		

$p > .05$.

Ho 6.6: There are no differences in grade point averages among achievement groups and effectiveness groups on difficult items.

The mean grade point averages were calculated by directions groups and by effectiveness groups on difficult items. These are presented in Table 68. The grade point averages were higher for ineffective and effective changers than for countereffective changers.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the change effectiveness groups on difficult items; F (4, 120 df) = .90, $p > .05$. There was no main effect for directions; F (2, 120 df) = 1.32, $p > .05$. There was a main effect for effectiveness of changes; F (2, 120 df) = 4.26, $p < .05$. (See Table 69.) The grade point averages were associated with the degree of change effectiveness on difficult items.

Table 68.--Mean grade point averages (GPA) and standard deviations (SD) for degree of change effectiveness on difficult items.

Group	Mean GPA-- Difficult Countereffective			Mean GPA-- Difficult Ineffective			Mean GPA-- Difficult Effective			Total Mean GPA-- Difficult		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	2.51	.62	3	3.08	.28	11	2.86	1.02	45	2.64	.74
Treatment 1	20	2.19	.72	5	3.05	.90	15	2.63	.88	40	2.46	.88
Treatment 2	25	2.39	.52	7	2.38	.95	11	2.78	.54	44	2.48	.62

Table 69.--ANOVA table for grade point average--difficult.

Source	df	MS	F
Treatment group	2	72.05	1.32
Degree of effectiveness	2	231.85	4.26*
Treatment x Degree	4	49.00	.90
Error	120	54.37	
Total	128	56.75	

*$p < .05$.

In summary, student achievement, as defined by scores on the objective final examination and grade point averages, was associated with the degree of change effectiveness on difficult items. However, achievement measures were not associated with directions and the degree of change effectiveness on total and easy items. Subjects who had higher grade point averages and higher objective final examination scores had a higher percentage of correct changes.

Ability and Answer Changing

Ho 7.1: There are no differences in SAT verbal scores among direction groups and effectiveness groups on total items.

The mean SAT verbal scores by directions groups and by effectiveness groups on total items were computed. These means are presented in Table 70. The mean SAT scores increased with the degree of change effectiveness.

Table 70.--Means and standard deviations (SD) for SAT verbal subtest scores for degree of change effectiveness on total items.

Group	Mean SAT Verbal--Total Countereffective			Mean SAT Verbal--Total Ineffective			Mean SAT Verbal--Total Effective			Total Mean SAT Verbal--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	469.13	71.21	5	508.00	76.61	17	515.29	67.27	45	490.88	72.29
Treatment 1	19	483.68	109.45	7	518.57	101.55	13	490.76	110.18	39	492.30	106.31
Treatment 2	11	452.72	90.34	11	455.45	86.18	22	460.00	104.41	44	457.04	90.87

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the effectiveness groups; F (4, 119 df) = .35, $p \succ .05$. There was no main effect for directions; F (2, 119 df) = 2.71, $p > .05$). There was no main effect for effectiveness of changes; F (2, 119 df) = 1.05, $p > .05$. (See Table 71.) The SAT verbal scores were not associated with the directions or with the degree of change effectiveness.

Table 71.--ANOVA table for mean SAT verbal--total.

Source	df	MS	F
Treatment group	2	22452.56	2.71
Degree of effectiveness	2	8713.95	1.05
Treatment x Degree	4	2936.89	.36
Error	119	8280.64	
Total	127	8259.74	

$p > .05$.

Ho 7.2: There are no differences in SAT verbal scores among direction groups and effectiveness groups on easy items.

The SAT verbal score means by directions groups and by effectiveness groups on easy items were calculated. These are presented in Table 72. Again, these scores gradually improved with the degree of change effectiveness.

Table 72.--Mean SAT verbal scores and standard deviations (SD) for degree of change effectiveness on easy items.

Group	Mean SAT Verbal--Easy Countereffective			Mean SAT Verbal--Easy Ineffective			Mean SAT Verbal--Easy Effective			Total Mean SAT Verbal--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	27	483.33	82.60	4	487.50	32.01	14	506.42	58.52	45	490.88	72.29
Treatment 1	23	468.69	92.84	4	570.00	21.60	11	500.90	131.64	38	488.68	104.11
Treatment 2	15	457.33	105.45	3	406.66	72.34	26	442.69	84.97	44	457.04	90.87

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups and the effectiveness groups; F (4, 118 df) = 1.13, $p > .05$. There was no main effect for directions; F (2, 118 df) = 2.40, $p > .05$. There was no main effect for effectiveness of changes; F (2, 118 df) = .92, $p > .05$. (See Table 73.) The SAT verbal scores were not associated with directions or the degree of change on easy items.

Table 73.--ANOVA table for SAT verbal--easy.

Source	df	MS	F
Treatment group	2	19516.62	2.46
Degree of effectiveness	2	7358.90	.93
Treatment x Degree	4	8994.25	1.13
Error	118	7932.37	
Total	126	8077.90	

$p > .05$.

Ho 7.3: There are no differences in SAT verbal scores among direction groups and effectiveness groups on difficult items.

The mean SAT verbal scores by directions groups and by effectiveness groups were computed. These means are presented in Table 74. The mean SAT scores improved with the degree of change effectiveness.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There was no interaction between the directions and the effectiveness of changes. There was no main effect for directions;

Table 74.--Mean SAT verbal scores and standard deviations (SD) for degree of change effectiveness on difficult items.

Group	Mean SAT Verbal--Difficult Countereffective			Mean SAT Verbal--Difficult Ineffective			Mean SAT Verbal--Difficult Effective			Total Mean SAT Verbal--Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	474.8	64.90	3	540.0	104.40	11	522.7	74.40	45	490.8	72.20
Treatment 1	20	448.5	85.80	5	540.0	86.60	15	536.0	106.80	40	493.0	105.00
Treatment 2	26	441.5	85.80	7	461.4	98.00	11	490.9	97.10	44	457.0	90.80

F (4, 120 df) = .59, $p > .05$. There was a main effect for effectiveness; F (2, 120 df) = 7.20, $p < .01$. The SAT verbal scores were associated with the degree of change effectiveness on difficult items. (See Table 75.) Students who were more effective changers on difficult items had higher SAT verbal scores. The increase of the mean SAT verbal scores with the degree of change effectiveness was sufficient enough to be significant.

Table 75.--ANOVA table for SAT verbal--difficult.

Source	df	MS	F
Treatment group	2	18168.42	2.44
Degree of effectiveness	2	53439.40	7.20*
Treatment x Degree	4	4436.44	.60
Error	120	7422.46	
Total	128	8207.81	

*p = .001.

Ho 7.4: There are no differences in SAT mathematics scores among direction groups and effectiveness groups on total items.

The mean SAT mathematics scores by directions groups and by effectiveness groups on total items were calculated. (See Table 76.) The ineffective changers in each group had a higher mean SAT mathematics score than the effective changers.

The analysis of variance failed to reject the null hypothesis ($p > .05$). There were no interactions between the directions groups

Table 76.--SAT mathematics (MATH) mean and standard deviation (SD) scores for degree of change effectiveness on total items.

Group	Mean SAT MATH-- Total Countereffective			Mean SAT MATH-- Total Ineffective			Mean SAT MATH-- Total Effective			Total Mean SAT MATH-- Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	23	486.08	74.05	5	556.00	45.60	17	488.20	116.63	45	494.66	91.71
Treatment 1	19	475.26	96.34	7	510.00	100.10	13	501.50	102.62	39	490.25	97.64
Treatment 2	11	474.54	119.08	11	484.54	118.69	22	490.00	108.40	44	484.77	109.60

and the effectiveness groups; F (4, 119 df) = .36, p > .05. There was no main effect for directions; F (2, 119 df) = .27, p > .05. There was no main effect for effectiveness; F (2, 119 df) = .80, p > .05. The SAT mathematics scores were not associated with directions or the degree of change effectiveness on total items. (See Table 77.)

Table 77.--ANOVA table for SAT mathematics--total.

Source	df	MS	F
Treatment group	2	28.15	.27
Degree of effectiveness	2	82.65	.80
Treatment x Degree	4	37.68	.37
Error	119	102.36	
Total	127	98.57	

p > .05.

Ho 7.5: There are no differences in SAT mathematics scores among direction groups and effectiveness groups on easy items.

The means for SAT mathematics scores by directions groups and by effectiveness groups on easy items were calculated. These are presented in Table 78. The mean scores for the strong-directions group (treatment 2) gradually increased, but the mean scores for the other two groups were not consistent.

The analysis of variance failed to reject the null hypothesis (p > .05). There were no interactions between directions groups and effectiveness groups on easy items; F (4, 118 df) = 1.64, p > .05.

Table 78.--SAT mathematics (MATH) mean and standard deviation (SD) scores for degree of change effectiveness on easy items.

Group	Mean SAT MATH-- Easy Countereffective			Mean SAT MATH-- Easy Ineffective			Mean SAT MATH-- Easy Effective			Total Mean SAT MATH-- Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	27	502.96	122.69	4	465.00	86.60	14	487.14	74.87	45	494.60	91.71
Treatment 1	23	463.40	121.57	4	582.50	39.47	11	500.00	80.20	38	486.57	96.57
Treatment 2	15	504.00	116.60	3	450.00	115.30	26	477.69	107.60	44	484.77	109.68

There was no main effect for directions; F (2, 118 df) = .10, p > .05. There was no main effect for change effectiveness; F (2, 118 df) = .14, p > .05. (See Table 79.) The SAT mathematics scores were not associated with directions or with change effectiveness on easy items.

Table 79.--ANOVA table for SAT mathematics--easy.

Source	df	MS	F
Treatment group	2	10.53	.13
Degree of effectiveness	2	13.95	.14
Treatment x Degree	4	162.29	1.64
Error	118	98.70	
Total	126	98.01	

p > .05.

Ho 7.6: There are no differences in SAT mathematics scores among direction groups and effectiveness groups on difficult items.

The means and standard deviations for SAT mathematics subtests were calculated by directions groups and by effectiveness groups on difficult items. These are presented in Table 80. The mean SAT mathematics scores gradually increased with the level of change effectiveness. Again, the ineffective changers scored the highest mean.

The analysis of variance failed to reject the null hypothesis (p > .05). There were no interactions between the directions groups and the effectiveness-of-change groups; F (4, 120 df) = .64, p > .05. There was no main effect for directions; F (2, 120 df) = .42, p > .05.

Table 80.--SAT mathematics (MATH) mean and standard deviation (SD) scores for degree of change effectiveness on difficult items.

Group	Mean SAT MATH-- Difficult Countereffective			Mean SAT MATH-- Difficult Ineffective			Mean SAT MATH-- Difficult Effective			Total Mean SAT MATH-- Difficult		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control	31	477.70	80.50	3	586.00	152.77	11	517.20	116.20	45	494.66	91.70
Treatment 1	20	446.00	76.30	5	520.00	109.00	15	538.60	94.50	40	490.00	96.30
Treatment 2	26	463.40	107.10	7	500.00	103.40	11	525.40	115.80	44	484.70	109.60

There was a main effect for degree of change effectiveness; F (2, 120 df) = 6.91, $p < .01$. The SAT mathematics scores were associated with the degree of change effectiveness. (See Table 81.) The students who were more effective changers had higher SAT mathematics scores.

Table 81.--ANOVA table for SAT mathematics--difficult.

Source	df	MS	F
Treatment group	2	38.77	.42
Degree of effectiveness	2	633.95	6.91*
Treatment x Degree	4	58.71	.64
Error	120	91.628	
Total	128	97.81	

*p = .001.

In summary, SAT verbal and mathematics subtest scores were associated with the effectiveness of changes on difficult items. Effective answer changers on difficult items had higher mean mathematics and verbal test scores than countereffective changers, but ineffective changers had the highest mean SAT scores. The same relationships were not found for change effectiveness on total items or easy items.

Summary

The major findings reported within the limitations of the study in this chapter were:

1. The directions to change answers did not affect the number of answers changed.

2. The directions to change answers affected the percentage of correct changes on easy items. However, that effect was not maintained when difficult items or easy items were considered.

3. The SSHA work methods scores were associated with the effectiveness of changes on difficult items. The SSHA subtests, education acceptance and teacher approval, were associated with answer changing effectiveness on total items. The SSHA composite score, study attitudes, was also associated with the answer changing effectiveness on total items, as was the SSHA total score, study orientation. The SSHA scores were not associated with effective answer changing on easy items. The SSHA subtest delay avoidance and effective answer changing were not associated with answer changing effectiveness on total, easy, or difficult items.

4. Item difficulty was not related to effective answer changing.

5. Neither males nor females were more effective answer changers or were more affected by directions.

6. Student achievement, as measured by the objective final scores and the subjects' grade point averages, was associated with the answer changing effectiveness on difficult items. Student achievement was not associated with answer changing effectiveness on easy or total items or by directions to change answers.

7. Student ability, as measured by the SAT verbal and mathematics subtest scores, was associated with the degree of change effectiveness

on difficult items. Student ability was not associated with the degree of change effectiveness on easy items or on total items.

In Chapter V, a summary of the study, a discussion of the data, and recommendations for future study are presented.

CHAPTER V

SUMMARY, DISCUSSION, AND RECOMMENDATIONS

Summary

The purpose of this study was to investigate the effects of three different sets of directions to change answers on the answer changing behavior of college students who were taking an objective final examination in a required religion course in a liberal arts college in the northeastern United States in fall 1982. The study also investigated the differences in the Survey of Study Habits and Attitudes (SSHA) scores among different direction groups and among different change effectiveness groups. Also considered were differences in student achievement and ability among direction groups and change effectiveness groups. The relationship between item difficulty and answer changing was also investigated.

Major findings were that the directions to change answers affected answer changing behavior on easy items. The SSHA work methods score was associated with the effective changing of difficult items. The study attitudes subtests and study orientation total SSHA scores were associated with effective changing on total items.

Grade point averages, objective test scores, and SAT verbal and mathematics subtest scores were associated with effective answer changing on difficult items. No ability measures and no SSHA subtest

scores were associated with effective answer changing on easy items or on total items. Neither males nor females were more effective answer changers, nor were there relationships between the level of item difficulty and effective answer changing. (See Table 82.)

Limitations

Discussion of the data cannot be undertaken without considering the many limitations of the study. When the investigator chose to collect these data on this population, the decision was made with some awareness of the potential difficulties: possible faking of responses on the SSHA, a lengthy final examination that might inhibit the perceived opportunity for answer revision, the unknown level of difficulty of the objective final exam, and the possibility that students would not read the directions concerning changing responses.

Students did find the test lengthy. Many stayed past the hour of the examination to complete the test. The difficulty level of the final examination as measured by the standard deviation (7.52) did reveal the test to be somewhat easy. In addition, there was some concern that some items did not discriminate between the high-scoring and low-scoring subjects. (See Appendix C.) Subjects blotted out approximately 8 percent of their original responses before marking the final answers, suggesting that some subjects might not have read the directions. Although some researchers have assigned such revisions at random and have used a wrong-to-wrong category, this investigator chose to assign the blotted wrong answers to one of two categories: right-to-wrong or wrong-to-right.

Table 82.--Summary of research findings for null hypotheses.

	Null Hypotheses	Findings
Ho 1	There is no effect due to directions on the number of answers changed.	Failure to reject
Ho 2.1	There is no effect due to directions on the percentage of correct changes on total items.	Failure to reject
Ho 2.2	There is no effect due to directions on the percentage of correct changes on easy items.	Rejected, p = .006
Ho 2.3	There is no effect due to directions on the percentage of correct changes on difficult items.	Failure to reject
Ho 3.1	There are no differences in SSHA delay avoidance scores among direction groups and effectiveness groups on total items.	Failure to reject
Ho 3.2	There are no differences in SSHA delay avoidance scores among direction groups and effectiveness groups on easy items.	Failure to reject
Ho 3.3	There are no differences in SSHA delay avoidance scores among direction groups and effectiveness groups on difficult items.	Failure to reject
Ho 3.4	There are no differences in SSHA work methods scores among direction groups and effectiveness groups on total items.	Failure to reject
Ho 3.5	There are no differences in SSHA work methods scores among direction groups and effectiveness groups on easy items.	Failure to reject
Ho 3.6	There are no differences in SSHA work methods scores due to effectiveness of answer changing on difficult items.	Rejected, p = .048
Ho 3.7	There are no differences in SSHA education acceptance scores due to effectiveness of answer changing on total items.	Rejected, p = .01

Table 82.--Continued.

	Null Hypotheses	Findings
Ho 3.8	There are no differences in SSHA education acceptance scores among direction groups and effectiveness groups on easy items.	Failure to reject
Ho 3.9	There are no differences in SSHA education acceptance scores among direction groups and effectiveness groups on difficult items.	Failure to reject
Ho 3.10	There are no differences in SSHA teacher approval scores due to effectiveness of answer changing on total items.	Rejected, $p = .032$
Ho 3.11	There are no differences in SSHA teacher approval scores among direction groups and effectiveness groups on easy items.	Failure to reject
Ho 3.12	There are no differences in SSHA teacher approval scores among direction groups and effectiveness groups on difficult items.	Failure to reject
Ho 3.13	There are no differences in SSHA composite score study habits (delay avoidance and work methods) among direction groups and effectiveness groups on total items.	Failure to reject
Ho 3.14	There are no differences in SSHA composite score study habits (delay avoidance and work methods) among direction groups and effectiveness groups on easy items.	Failure to reject
Ho 3.15	There are no differences in SSHA composite score study habits (delay avoidance and work methods) among direction groups and effectiveness groups on difficult items.	Failure to reject
Ho 3.16	There are no differences in SSHA study attitudes scores due to effectiveness of answer changing on total items.	Rejected, $p = .023$

Table 82.--Continued.

	Null Hypotheses	Findings
Ho 3.17	There are no differences in SSHA composite score study attitudes (education acceptance and teacher approval) among direction groups and effectiveness groups on easy items.	Failure to reject
Ho 3.18	There are no differences in SSHA composite score study attitudes (education acceptance and teacher approval) among direction groups and effectiveness groups on difficult items.	Failure to reject
Ho 3.19	There are no differences in SSHA study orientation scores due to effectiveness of answer changing on total items.	Rejected, $p = .039$
Ho 3.20	There are no differences in SSHA study orientation scores among direction groups and effectiveness groups on easy items.	Failure to reject
Ho 3.21	There are no differences in SSHA study orientation scores among direction groups and effectiveness groups on difficult items.	Failure to reject
Ho 4	There is no relationship between item difficulty and effective answer changing.	Failure to reject
Ho 5.1	There are no differences in percentage of correct changes among direction groups and sex groups on total items.	Failure to reject
Ho 5.2	There are no differences in percentage of correct changes among direction groups and sex groups on easy items.	Failure to reject
Ho 5.3	There are no differences in percentage of correct changes among direction groups and sex groups on difficult items.	Failure to reject
Ho 6.1	There are no differences in the objective test scores among direction groups and effectiveness groups on total items.	Failure to reject
Ho 6.2	There are no differences in the objective test scores among direction groups and effectiveness groups on easy items.	Failure to reject

Table 82.--Continued.

	Null Hypothesis	Findings
Ho 6.3	There are no differences in objective test scores due to effectiveness of answer changing on difficult items.	Rejected, p = .01
Ho 6.4	There are no differences in the grade point averages among direction groups and effectiveness groups on total items.	Failure to reject
Ho 6.5	There are no differences in the grade point averages among direction groups and effectiveness groups on easy items.	Failure to reject
Ho 6.6	There are no differences in grade point averages due to effectiveness of answer changing on difficult items.	Rejected, p = .016
Ho 7.1	There are no differences in the SAT verbal scores among directions groups and effectiveness groups on total items.	Failure to reject
Ho 7.2	There are no differences in the SAT verbal scores among directions groups and effectiveness groups on easy items.	Failure to reject
Ho 7.3	There are no differences in the SAT verbal scores due to effectiveness of answer changing on difficult items.	Rejected, p = .001
Ho 7.4	There are no differences in the SAT mathematics scores among direction groups and effectiveness groups on total items.	Failure to reject
Ho 7.5	There are no differences in the SAT mathematics scores among direction groups and effectiveness groups on easy items.	Failure to reject
Ho 7.6	There are no differences in SAT mathematics scores due to effectiveness of answer changing on difficult items.	Rejected, p = .001

The SSHA was administered the last class period before the final examination. It was thought that few students would miss this class. Contrary to expectations, approximately 40 subjects out of 147 who were still enrolled in the class did not take the SSHA during class. This meant that the SSHA had to be administered at the time of the final examination. This had some potential problems. First, it made the total test time 35 minutes longer for some subjects. This might have inhibited their changing patterns. It also might have encouraged subjects to rush through the SSHA, giving less accurate or meaningful self-reports. Several subjects were dropped from the study due to lack of completion of the SSHA. They accounted for a portion of the 18 subjects who were eliminated because of insufficient data. It should be noted that more students were lost to the study because of missing SAT scores than missing SSHA scores. Finally, taking the SSHA on the day of the final meant that the investigator was present at the time of the final examination. This might have alerted the subjects to the potential existence of an experiment. Although a control group was used in the design, this still could have affected the patterns of changing.

The last limitation of this study has to do with the missing SAT scores on transfer students and the subsequent exclusion of that group from the study. The effects of this on the study are really not known. The data concerning the equivalency of the groups showed that all groups were equal. (See Appendix D.) However, treatment 2 did have SAT verbal scores of 457 as compared with 490 for the control and 492 for treatment 1. When the effect on the achievement and ability

measures of answer changing effectiveness for higher achieving students is considered, the lower SAT scores for treatment 2 might have influenced the results of the study.

Discussion

Analysis of the data indicated that the directions to change answers given to the three groups did not affect the number of changes made. There was, however, a higher mean number of changes in the strong-directions-to-change group (treatment 2) than in the caution group (treatment 1) or the no-specific-directions group (control). These differences were not significant. (See Tables 1 and 2, p. 47.)

The subjects in this study changed an average of 2.5 answers. Mercer (1978), in a review of 23 studies, noted a range of 1.1 to 17.0 mean changes per student and a median of 4.4 changes per student. In studies where directions to change have been incorporated into the design of the study, higher mean changes have been noted. Jacobs (1972) reported 14.7 mean changes per student, Foote and Belinky (1972) reported 8.8 and 4.7 mean changes, Smith and Moore (1976) reported 13.2 and 17.0 mean changes, and Sutton (1982) reported 5.2 mean changes. Subjects in the present study changed fewer responses than would have been expected. The range of changes in the present study was 0-12 changes.

Examination of the raw data indicated that 93 percent of the subjects in the present study changed responses. Mercer (1978) noted a range of 61-100 percent of subjects changing some responses

in her review. Mueller and Wasser (1977) noted a median of 89 percent of the students changing responses in the seven studies they reviewed. Findings concerning percentage of students changing answers appear to be in keeping with previous research.

Of the 20 subjects who did not revise any answers in the present study, 10 were in the no-specific-directions group (22 percent of the control group), 6 were in the caution group (15 percent of treatment 1), and 4 were in the strong-direction group (9 percent of treatment 2). Because data were unavailable to ascertain these subjects' behavior on previous testing, the findings of Stoffer, Davis, and Brown (1977)--that some students appear to have a no-change response-set--cannot be confirmed.

Jarrett (1947) noted that some students change a larger proportion of answers than others. In this study, there were five subjects in the control group who accounted for 35 percent of the answer changing. Three subjects changed 27 percent of the total responses in the treatment 1 group, whereas three subjects changed 16 percent of the responses in the treatment 2 group. In the no-specific-change group the mode for change was 1, in the caution group the mode was 1, and in the strong-direction group the mode was 2. (See Appendix E.)

Both the no-specific-directions group and the caution group had more subjects who changed no answers and subjects who accounted for a greater percentage of changes. The mode was higher for the strong-direction group. Although the mean changes were not significantly different among the groups, an examination of the data indicated that there were some different patterns within the groups.

One explanation for the possible lower rate of changes in the present study is the difficulty level of the objective final on which the answer changing was observed. The examination had a mean of 55.84 and a standard deviation of 7.52. The examination might not have been difficult enough to encourage much revision. Another possible explanation was that the test was lengthy enough that students did not perceive that they had the time to revise their answers.

Because the test was thought to be too easy to provoke a higher rate of answer changing, changes were examined on easy items and difficult items as well as total items. The thought was that patterns of answer changing might differ on easy and difficult items.

Analysis of the data concerning directions to change answers and the degree of change effectiveness indicated that strong directions to change answers significantly affected answer changing on easy items. Students who were told to go back and revise answers profited from revision of easy items. Analysis of data also indicated that there were no benefits for the strong-direction group on difficult or total items. One possible explanation for this finding could be that when the students were instructed to go back and revise answers, they found and corrected their clerical errors on easy items. And they found and corrected errors from misreading questions and perhaps gained points from information from other items. The answers that were more difficult were not revised as effectively because subjects simply did not know the correct answers. Students in the control and treatment 1 groups might have specifically gone back and checked selectively those answers that were difficult for

them. Perhaps this caused them to miss the opportunity to revise their clerical errors and errors caused by misreading the easier questions or to pick up points on the easier items by using knowledge they gained on subsequent questions. Sutton (1982) found more benefit for his adult subjects from such clerical types of revisions.

Table 83 in Appendix F presents data concerning the number of students who gained, lost, or were not affected by changes. Table 84 presents data concerning mean positive and negative changes by group.

Subjects in all three treatment groups who had higher work methods SSHA scores were the superior answer changers on difficult items. A possible explanation was that those with more self-reported study skill knew more and could therefore revise answers on difficult items more easily. There were no other significant effects on SSHA scores by effective changing of difficult items. Apparently, student self-reported time-management strategies and attitudes toward teachers and education were not associated with changing difficult answers effectively. It should be noted that there were no interactions among the work methods score, the degree of change effectiveness, and the directions to change answers.

Answer changing effectiveness on total items was associated with education acceptance, teacher approval, and the composite study attitudes SSHA scores. Students with better attitudes toward schooling did a more effective job of changing answers than those who did not express positive attitudes toward schooling. One explanation for this finding is that those with more positive attitudes possibly found it easier to spend more time attempting to

go back and proofread their examinations. The assumption here is that the subjects believed that answer revision would increase their scores. The finding concerning attitude in the present study supported the findings of McMorris, Lichtstein, and Hoops (1983), who reported a relationship between a positive attitude toward the course and course tests and effective answer changing. A similar finding was reported by Sitton, Adams, and Anderson (1980).

The total score for the SSHA, study orientation, was associated with the degree of change effectiveness on total items. Countereffective changers did have the lowest study orientation scores. (See Table 45, p. 89.) The highest scores on study orientation, however, were held by those subjects who were ineffective changers. Examination of the data revealed that the number of students in this ineffective group was not great enough to generate meaningful implications. This phenomenon, however, suggested a need for further study. Possibly there was a group of students who had studied well, and when they attempted to change answers they read too much into the questions and changed right answers to wrong ones. This group could also have reached the "ceiling" effect on the examination, and their changes could, therefore, have been from wrong-to-wrong. Even though the ineffective changers had the highest mean study orientation scores, the study orientation scores were significantly higher for the effective changers than for countereffective changers. There was no interaction with the treatment. (See Table 44, p. 88.)

It should be noted that the SSHA scores were not associated with effective answer changing on easy items. This finding was of

particular interest because strong directions to change answers did significantly affect answer changing on easy items. High scores on self-reported study habits and attitudes were not associated with effective changes on easy items. This finding suggested a need for further research into the reasons that students revised answers. Were the revisions made by the strong-directions group a result of revision of clerical errors? If this is true, it appears that directions to revise answers tended to help some students to be test-wise--to make use of the format of the test and the testing situation to improve their scores.

Neither males nor females were more effective answer changers. This finding replicated the finding of Mueller and Wasser (1977), who reported that sex differences were usually present in studies where one sex or the other had a higher mean test score. In this study, there were no differences between the mean test scores of males or females. Data analysis also indicated that there was no difference in effective answer changing on difficult or on easy items. This finding replicated the finding of Ballance (1975).

Data analysis indicated that objective test scores, grade point averages, and SAT scores were associated with the degree of effective answer changing on difficult items. Achievement and ability measures were not associated with effectiveness of changes on easy or total items. The finding that higher-achieving students were more effective answer changers may indicate that higher-ability or higher-achieving students are more likely to be more effective answer changers on the more difficult test questions, but their effectiveness is not superior

on the easier item changes. Jacobs (1972) and Vidler and Hansen (1980) did not find a relationship between effective changing on difficult items and high achievement. Others who have investigated achievement and ability measures and effectiveness of answer revision have not found relationships between these variables when total item changes have been studied (McMorris & Leonard, 1976; Pascale, 1974; Reiling & Taylor, 1972; Stoffer et al., 1977). This study replicated these findings.

Implications of the Study

Within the limitations of population and methodology of this study, implications are suggested for educators. Educators should encourage students to go back to check their answers on objective tests. Answers that have been incorrectly marked, misread, or that now appear incorrect should be revised. Such revisions will probably improve scores.

A college instructor of a study skills course might include the following information concerning answer changing: "Go back and revise your answers; you will improve your scores. However, the improvement will probably be from the changes that you have made on easy items. You will not be likely to make as many points from revision of difficult items if you do not have good study skills and if you do not usually score well on classroom tests and standardized tests, such as the SAT. If you go through the test a second time, particularly be aware of the items on which you feel you have made a clerical error, you think you have misread, and items where you now know the

answer because a subsequent item gave you a clue about the answer. These will most likely be the easier test items. Carefully consider changing items on which you are still not sure between two answers. These are probably the more difficult items, and your chances of changing them from a wrong to a right answer are more related to your level of achievement and study ability."

Recommendations for Future Research

The findings of the present study justify further investigation into the effects of directions to change answers on answer changing behavior of students.

1. Replication of the design of the present study using a course test that is not as lengthy and that has a higher difficulty and discrimination level is suggested. Further, replication with a younger population is also needed. Such research might attempt to test for the developmental stage at which children are capable of reversing answers (Piaget, 1972). Replication of the finding for effective revision of easy items is needed.

2. Investigation into the reasons students revise answers is also suggested. Interviewing students immediately after an examination for explanations of their revisions might give insight into the present finding concerning the effectiveness of students on easy items in the strong-directions groups.

3. Investigation into the reasons that students in the ineffective answer changing category achieved high scores on different achievement and ability measures but lost points changing is also

suggested. The small number of ineffective changers in this study inhibits the report of meaningful implications.

4. Investigation into the students' interpretation of the words "change with caution" is needed. In this study, the behavior of subjects in the caution group was more similar to subjects in the no-specific-directions group than it was similar to the strong-directions group. Perhaps "change with caution" in a testing situation has a different connotation to students than it does to the instructors and study-skills-textbook writers.

5. Investigation into personality factors associated with nonchangers, countereffective changers, and changers who change larger proportions of answers is also needed.

APPENDICES

APPENDIX A

LETTER TO STUDENTS WHO TOOK THE SSHA WITH THE FINAL EXAMINATION

Dear Student,

Because you were not present in class on Monday to complete the Survey of Study Habits and Attitudes, you need to do that now. The retention committee of Gordon College is interested in helping you do well at Gordon.

Instructions:

1. Use the special pencil provided.
2. Fill in the information on the answer sheet. On the answer sheet, print your name, your age, your sex (M or F), and the date, which is 11/10/82. The Form is Form C. Now circle your year in school.
3. Now read the directions in the answer booklet on the first page and begin working.

Turn in the Survey of Study Habits and Attitudes booklet, answer sheet, and pencil with your final examination.

Thank you,

Pat Wilson

PW:jr

APPENDIX B

TEST DIRECTIONS TO SUBJECTS

CONTROL

Name:

I.D.:

Bible 101 UNDERSTANDING THE OLD TESTAMENT

Final Examination: Part B, November 1982

This is Part B of your final examination. It consists of 40 multiple-choice questions and 40 true or false statements, worth one point each. This section will account for 20% of your final course grade.

INSTRUCTIONS

Please read the instructions carefully, and answer the questions asked. If you find that you have made an error, cross out (X) the wrong answer and write another answer.

TREATMENT 1

Name:

I.D.:

Bible 101 UNDERSTANDING THE OLD TESTAMENT

Final Examination: Part B, November 1982

This is Part B of your final examination. It consists of 40 multiple-choice questions and 40 true or false statements, worth one point each. This section will account for 20% of your final course grade.

INSTRUCTIONS

If you feel that you have made an error, cross out (X) the answer you believe is wrong and write another answer. Change with caution answers you believe are incorrect.

TREATMENT 2

Name:

I.D.:

Bible 101 UNDERSTANDING THE OLD TESTAMENT

Final Examination: Part B, November 1982

This is Part B of your final examination. It consists of 40 multiple-choice questions and 40 true or false statements, worth one point each. This section will account for 20% of your final course grade.

INSTRUCTIONS

If you find that you have made an error, cross out (X) the answer you believe is wrong and mark another answer. After you have worked through the test once, go back and change the answers you now believe to be incorrect. Research indicates that revising answers will improve your score on this test.

GO BACK AND REVISE YOUR ANSWERS!!

APPENDIX C

ITEM DIFFICULTY AND DISCRIMINATION DATA FOR THE OBJECTIVE TEST

Discrimination Data

Item	Difficulty	Discrimination	Item	Difficulty	Discrimination
1	12	11	41	60	28
2	75	23	42	9	8
3	61	25	43	23	37
4	71	32	44	22	-12
5	23	28	45	59	6
6	42	49	46	18	31
7	15	20	47	32	54
8	20	34	48	10	8
9	29	51	49	48	28
10	27	38	50	10	0
11	7	17	51	21	15
12	omitted		52	7	23
13	21	22	53	12	20
14	32	49	54	48	-48
15	13	3	55	30	38
16	18	23	56	23	26
17	10	23	57	10	17
18	0	0	58	95	- 6
19	5	14	59	60	- 3
20	68	43	60	12	3
21	20	51	61	24	0
22	23	48	62	4	6
23	12	8	63	72	- 9
24	72	54	64	19	40
25	36	40	65	18	17
26	32	6	66	11	11
27	32	23	67	62	20
28	11	20	68	48	40
29	24	23	69	8	14
30	66	63	70	72	31
31	65	62	71	12	23
32	14	28	72	15	17
33	52	26	73	28	54
34	13	31	74	2	6
35	15	34	75	18	34
36	35	40	76	60	25
37	30	51	77	18	34
38	33	46	78	17	- 6
39	10	17	79	35	12
40	12	3	80	12	22

APPENDIX D

GROUP EQUIVALENCY DATA

Group Equivalency for SAT--Verbal

Source	df	MS	F
Between groups	2	17207.41	2.12
Within groups	125	8116.58	
Total	127		

p = .12

Group Equivalency for SAT--Mathematics

Source	df	MS	F
Between groups	2	10.92	.10
Within groups	125	99.97	
Total	127		

p = .89

Group Equivalency for Study Orientation

Source	df	MS	F
Between groups	2	215.68	.22
Within groups	126	965.27	
Total	128		

p = .80

Group Equivalency for Objective Final Scores

Source	df	MS	F
Between groups	2	38.21	.67
Within groups	126	57.03	
Total	128		

p = .51

Group Equivalency for Grade Point Averages

Source	df	MS	F
Between groups	2	39.71	.69
Within groups	126	57.02	
Total	128		

p = .50

APPENDIX E

NUMBER OF CHANGES PER SUBJECT BY GROUP

Control Group (N = 45)

Subject	Number of Changes
011	9
012	3
013	1
014	2
015	0
016	1
017	6
018	0
019	5
110	1
111	0
112	0
113	4
114	0
115	1
116	1
117	3
118	0
119	0
120	5
121	0
122	3
123	7
124	1
125	2
126	0
127	2
128	5
129	1
130	1
131	3
132	1
133	4
134	1
135	9
136	2
137	4
138	4
139	1
140	2
141	7
142	1
143	0
144	2
145	2

Treatment 1 (N = 40)

Subject	Number of Changes
021	2
022	1
023	4
024	3
025	3
026	12
027	4
028	1
029	6
210	0
211	4
212	1
213	0
214	1
215	0
216	3
217	0
218	5
219	3
220	5
221	2
222	8
223	1
224	3
225	0
226	1
227	3
228	3
229	1
230	3
231	2
232	2
233	1
234	2
235	1
236	1
237	0
238	2
239	2
240	2

Treatment 2 (N = 44)

Subject	Number of Changes
031	5
032	5
033	1
034	5
035	2
036	3
037	0
038	2
039	7
310	4
311	7
312	2
313	7
314	1
315	2
316	2
317	2
318	2
319	1
320	5
321	2
322	3
323	0
324	8
325	2
326	2
327	3
328	3
329	5
330	1
331	4
332	1
334	0
335	0
336	5
337	5
338	3
339	3
340	2
341	5
342	2
343	5
344	5

APPENDIX F

TABLES 83 AND 84

Table 83.--Number of students who gained points, lost points, or were not affected by changes (neutral), by group.

Group	N	Lost	Gained	Neutral
Control	45	7	16	22
Treatment 1	40	10	12	18
Treatment 2	44	13	19	12
Total	129	30	47	52

Student gain-to-loss ratio = 1.56:1.

Table 84.--Student means and standard deviations (SD) for positive (+) and negative (-) changes, by group.

Group	N	Mean + Changes	SD	Mean - Changes	SD
Control	45	1.31	1.66	1.55	1.41
Treatment 1	40	1.30	1.44	1.15	1.44
Treatment 2	44	1.90	1.44	1.18	1.22

REFERENCES

REFERENCES

Archer, N. S., & Pippert, R. (1962). Don't change the answer! Clearing House, 37, 39-41.

Armstrong, W. H. (1975). Study tips. Woodbury, NY: Barron's Education Series, Inc.

Ballance, C. T. (1975). An investigation of the relations of the alteration of responses to objective test questions to selected student and test item characteristics. Dissertation Abstracts International, 36, 5220A (University Microfilms No. 76-4592).

Ballance, C. T. (1977). Students' expectations and their answer-changing behavior. Psychological Reports, 41, 163-166.

Ballance, C. T., & Gentzel, W. E. (1980). Effects of several scoring techniques on answer-changing by students. Psychological Reports, 46, 653-654.

Bath, J. A. (1967). Answer-changing behavior on objective examinations. Journal of Educational Research, 61, 105-107.

Beck, M. D. (1978). The effect of item response changes on scores on an elementary reading achievement test. Journal of Educational Research, 71, 153-156.

Berriens, F. K. (1939). Are first impressions best on objective tests? School and Society, 50, 319-320.

Best, J. B. (1979). Item difficulty and answer changing. Teaching of Psychology, 6, 228-230.

Brown, W. F., & Holtzman, W. H. (1965). Survey of study habits and attitudes. New York: Psychological Corporation.

Brown, W. F., & Holtzman, W. H. (1967). Survey of study habits and attitudes: SSHA manual. New York: Psychological Corporation.

Clark, C. A. (1962). Should students change answers on objective tests? Chicago Schools Journal, 43, 382-385.

Copeland, D. A. (1972). Should chemistry students change answers on multiple-choice tests? Journal of Chemical Education, 49, 258.

Crocker, L., & Benson, J. (1980). Does answer-changing affect test quality? Measurement and Evaluation in Guidance, 12, 233-239.

Cromack, T. R. (1981, April). Some theoretical implications of changing answers on multiple-choice tests. Paper presented at the meeting of the New England Educational Research Organization, Lenox, MA.

Cummings, O. W. (1981). Impact of response changes on objective test characteristics and outcomes for junior high students. Measurement and Evaluation in Guidance, 14, 32-37.

Davis, R. E. (1975). Changing examination answers: An educational myth? Journal of Medical Education, 50, 685-687.

Deese, J. (1959). In O. K. Buros (Ed.), The fifth mental measurement yearbook reviews: Reading-study skills (p. 272). Highland, NJ: Gryphon Press.

Ebel, R. L. (1965). Measuring educational achievement. Englewood Cliffs, NJ: Prentice-Hall.

Ebel, R. L. (1972). Essentials of educational measurement (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Eysenck, H. J., & Eysenck, S. B. G. (1968). Eysenck personality inventory. San Diego: Educational and Industrial Testing Service.

Foote, R., & Belinky, C. (1972). It pays to switch? Consequences of changing answers on multiple-choice examinations. Psychological Reports, 31, 667-673.

Gilbert, H. W. (1982). Pathways: A guide to reading and study skills. Boston: Houghton-Mifflin.

Goldfried, G., & D'Zurilla, T. (1973). Prediction of academic competence by means of the Survey of Study Habits and Attitudes. Journal of Educational Psychology, 64, 116-122.

Greene, K. (1981). Item-response changes on multiple-choice tests as a function of test anxiety. Journal of Experimental Education, 49, 225-228.

Gronlund, N. E. (1967). Measurement and evaluation in teaching. New York: Macmillan.

Harnish, D. L., Hill, K. T., & Fyans, L. J. (1980, April). Development of a shorter, more reliable, and more valid measure of test motivation. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, MA.

Higgins, M. J. (1967). Test reviews. Journal of Counseling Psychology, 14, 392-393.

Hill, G. E. (1937). The effects of changed response in true-false tests. Journal of Educational Psychology, 28, 308-310.

Hills, J. R. (1981). Measurement and evaluation in the classroom. Columbus, OH: Charles E. Merrill.

Honig, F. (1973). Taking tests and scoring high. New York: Arco.

Jackson, P. F. (1978). Answer changing on objective tests. Journal of Educational Research, 71, 313-315.

Jacobs, S. S. (1971, February). An experimental analysis of answer changing behavior on objective tests. Paper presented at the annual meeting of the American Educational Research Association, New York.

Jarrett, R. E. (1948). The extra-chance nature of changes on students' response to objective test-items. Journal of General Psychology, 38, 243-250.

Jenkins, J. J. (1974). Remember that old theory of memory? Well, forget it! American Psychologist, 29, 785-795.

Johnston, J. J. (1975). Sticking with first responses on multiple-choice exams: For better or for worse? Teaching of Psychology, 2, 178-179.

Johnston, J. J. (1978). Answer-changing behavior and grades. Teaching of Psychology, 5, 44-45.

Juola, A. E. (1968). Examination skills and techniques. Lincoln, NE: Cliff's Notes, Inc.

Kellighan, T., Madaus, G. F., & Airasian, P. W. (1982). The effects of standardized testing. Boston: Kliewer-Nyhoff.

Lawson, E. E. (1935). What happens when the second judgement is recorded in a true-false test? Journal of Educational Psychology, 26, 223-227.

Lehman, H. C. (1928). Does it pay to change initial decisions in true-false tests? School and Society, 28, 456-458.

Lindeman, R. H., & Merenda, P. F. (1979). Educational measurement. Glenview, IL: Scott Foresman.

Lowe, M. L., & Crawford, C. C. (1929). First impression versus second thought in true-false tests. Journal of Educational Psychology, 20, 192-195.

Lyman, H. B. (1978). Test scores and what they mean. Englewood Cliffs, NJ: Prentice-Hall.

Lynch, D.)., & Smith, B. C. (1975). Item response changes: Effects on test scores. Measurement and Evaluation in Guidance, 7, 220-224.

Maiorana, V. P. (1980). How to study and learn in college. Englewood Cliffs, NJ: Prentice-Hall.

Mallinson, G. G., & Miller, D. J. (1956). The effects of second guessing in achievement scores of college tests. Thirteenth yearbook of the National Council on Measurement in Education, 24-26.

Matthews, C. O. (1929). Erroneous first impressions on objective tests. Journal of Educational Psychology, 20, 280-286.

McMorris, R. F. (1981, April). Answer changing and attitudes. Paper presented at the meeting of the American Educational Research Association, Los Angeles.

McMorris, R. F., & Leonard, G. (1976, April). Item response changes and cognitive style. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco.

McMorris, R. F., Lichtstein, S. F., & Hoops, A. J. (1983, April). Answer changing after instructions on answer changing. Paper presented at the conference of the American Educational Research Association, Montreal, Canada.

Mehrens, W. A., & Lehmann, I. J. (1978). Measurement and evaluation in education and psychology (2nd ed). New York: Holt, Rinehart & Winston.

Mercer, M. (1977). An update of 'Implications of changing answers on objective test items.' (ERIC Document Reproduction Service No. ED 163 051).

Mercer, M. (1978). Answer changing and students' test scores (Doctoral dissertation, Rutgers University, 1978). Dissertation Abstracts International, 40, 0214A.

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25, 707-726.

Millman, J., & Pauk, W. (1969). How to take tests. New York: McGraw-Hill.

Morris, F. L. (1961). The validity of the Brown-Holtzman Survey of Study Habits and Attitudes 1960 experimental revision for grades 7-12. Unpublished master's thesis, Wayne State University, Detroit, MI.

Mueller, D. J., & Shwedel, A. (1975). Some correlates of net gain resultant from answer changing on objective achievement test items. Journal of Educational Measurements, 12, 251-254.

Mueller, D. J., & Wasser, V. (1977). Implication of changing answers on objective test items. Journal of Educational Measurement, 14, 9-13.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. (1975). Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill.

Noll, V. H., Scannell, D. P., & Craig, R. C. (1979). Introduction to educational measurements. Boston: Houghton-Mifflin.

Norman, M. H., & Norman, E. S. K. (1976). How to read and study for success in college. New York: Holt, Rinehart & Winston.

Pascale, P. J. (1974). Changing initial answers on multiple-choice achievement tests. Measurement and Evaluation in Guidance, 6, 236-238.

Pauk, W. (1974). How to study in college (2nd ed.). Boston: Houghton-Mifflin.

Payne, B. D. (1983, April). Test anxiety, answer changing behavior, and race. Paper presented at the American Educational Research Association, Montreal, Canada.

Penfield, D., & Mercer, M. (1980). Answer changing and statistics. Educational Research Quarterly, 5, 50-57.

Piaget, J. (1972). The child's concept of physical causality. Potowa, NJ: Littlefield, Adams, & Co.

Range, L. M., Anderson, H. N., & Wesley, A. L. (1982). Personality correlates of multiple-choice answer-changing patterns. Psychological Reports, 51, 523-527.

Reile, P. J., & Briggs, L. J. (1952). Should students change their initial answers on objective-type tests? More evidence regarding an old problem. Journal of Educational Psychology, 43, 110-115.

Reiling, E., & Taylor, R. (1972). A new approach to the problem of changing initial responses to multiple choice questions. Journal of Educational Measurement, 9, 67-70.

Robinson, F. P. (1970). Effective study (4th ed.). New York: Harper & Row.

Sarnacki, R. E. (1979). Examination of test-wiseness in the cognitive test domain. Review of Educational Research, 9, 252-279.

Shay, C. B. (1972). In O. K. Burow (Ed.), The seventh mental measurement yearbook (pp. 782-783). Highland, NJ: Gryphon Press.

Shepherd, J. F. (1982). The Houghton-Mifflin study skills handbook. Boston: Houghton-Mifflin.

Sitton, L. R., Adams, I. G., & Anderson, H. N. (1980). Personality correlates of students' patterns of changing answers on multiple-choice tests. Psychological Reports, 47, 655-660.

Smith, A., & Moore, J. C. (1976). The effects of changing answers on scores of non-test-sophisticated examinees. Measurement and Evaluation in Guidance, 8, 252-254.

Smith, J. D. (1977). The relationship of content area and ability with multiple-choice test response alterations. Dissertation Abstracts International, 37, 5066A. (University Microfilms No. 77-3154).

Smith, M., White, K. P., & Coop, L. H. (1979). The effect of item type on the consequences of changing answers on multiple-choice tests. Journal of Educational Measurements, 16, 206-208.

Stanley, J. C., & Hopkins, K. D. (1972). Educational and psychological measurements and evaluation. Englewood Cliffs, NJ: Prentice-Hall.

Staton, T. F. (1982). How to study (7th ed.). Nashville, TN: How to Study.

Stoffer, G. R., Davis, K. E., & Brown, J. B., Jr. (1977). The consequences of changing initial answers on objective tests. A stable effect and a stable misconception. Journal of Educational Research, 70, 272-277.

Sutton, J. P., Jr. (1982). Influencing answer-changing behavior: Changing beliefs about the consequences of answer changing through consultation (Doctoral dissertation, Auburn University, 1982). Dissertation Abstracts International, 42, 5029A.

Thorndike, E. L. (1951). Reliability. In E. F. Linquist (Ed.), Educational measurement. Washington, D.C.: American Council on Education.

Vaas, C. E., & Nungester, R. J. (1982, March). An investigation of the answer changing behavior of adult learners. Paper presented at the meeting of the American Educational Research Association, New York.

Vidler, D., & Hansen, R. (1980). Answer changing on multiple choice tests. Journal of Experimental Education, 49, 18-20.

Wen, S., & Lui, A. (1976). The validity of each of the four subtests of the Survey of Study Habits and Attitudes for each of two samples of college students and under each of two treatment conditions involving use of released class time. Educational and Psychological Measurement, 36, 565-568.

Wilson, P. (1980). Answer changing behavior on objective tests: What is our responsibility? Epistle, 6, 27-31.

Wrenn, C. G. (1959). In O. K. Buros (Ed.), The fifth mental measurement yearbook reviews: Reading-study skills (pp. 272-273). Highland, NJ: Gryphon Press.