To the Graduate Council:

I am submitting herewith a thesis written by Kenneth Horner entitled "An Investigation of Reaction to Psychological Measurement." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Organizational Psychology.

John W. Lounsbury

John W. Lounsbury, Major Professor

We have read this thesis
and recommend its acceptance:

John M. Lawyer Jr.

Accepted for the Council:

L Evans Reed

Vice Chancellor
Graduate Studies and Research

Thesis
79
.H673
cop.2

AN INVESTIGATION OF REACTION TO PSYCHOLOGICAL MEASUREMENT

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Kenneth Horner

August 1979

1397056

DEDICATION

To my Mother and Father who have given me so much through their love and understanding, and to Larry, Debbie, Mike, Linda, and those other very special people whose friendship I shall always treasure. I would also like to dedicate this thesis to the memory of my aunt, Virginia Krevel, for the encouragement she provided.

ACKNOWLEDGMENTS

The author would like to express his gratitude to Dr. John W. Lounsbury without whose support and friendship this project could never have been completed. Thanks are due to Dr. John M. Larsen, Jr. whose insightful comments greatly improved this thesis and to Dr. Aaron Nurick whose unselfish help allowed me to meet the deadlines and finish this thesis. I would also like to express my appreciation to Dr. Michael Gordon for his helpful comments during the planning stages of this thesis. I am indebted to Ian Friedenberg whose earlier study provided the basis for this thesis, but more importantly I would like to express my appreciation to Ian, his wife, Debbie, and his daughter, Carrie, for their support and friendship during my first year at The University of Tennessee. Finally, special thanks are due Ms. Jean Robeson for typing this thesis.

ABSTRACT

The purpose of this study was to determine if a measurable relationship exists between subjective reactions to a test/testing situation and the reliability of the testing instrument. A review of the literature revealed a paucity of research dealing with the influence of subjective reactions on the psychometric properties of tests. The major emphasis in the literature concerned research dealing with test-anxiety. This literature was summarized in terms of the negative impact on performance of a number of variables associated with test-anxiety.

A questionnaire type instrument was presented to a sample (N = 233) of undergraduate university students using two instructional conditions. The instructional conditions consisted of a high reaction set of instructions designed to elicit more negative reactions by having the administrator of the test act in a rude and aversive manner and a low reaction condition which was intended to elicit a more positive participant reaction. The questionnaire consisted of four sections: demographic items, subscales from the Edwards Personal Preference Schedule (EPPS), a Reactions Questionnaire, and the State-Trait Anxiety Inventory (STAI). The complete questionnaire contained a total of 189 items. The order of presentation of the four sections of the questionnaire was held constant through the study. The EPPS subscales used in this study were the order, affiliation, endurance, and aggression scales. The Reactions Questionnaire was a modified

version of the one developed by Friedenberg (1978) and was intended as a measure of subjective reactions to the EPPS in such areas as participant reaction to item content and the context of administration.

A factor analysis performed on the responses to the Reactions Questionnaire revealed five orthogonal reaction factors that accounted for 76.7 percent of the common variance. Factor scores were computed for each individual on each of the five reaction factors. These factor scores were then rank-ordered and were used as a basis for splitting the sample into separate "high" and "low" subgroups after dichotomizing and trichotomizing the sample.

Three hypotheses were developed and tested. The first hypothesis was that individuals experiencing more negative reactions (i.e., those in the "high reactions" subgroups) to the EPPS would show lower EPPS reliabilities than individuals experiencing more positive reactions to the EPPS (i.e., those in the "low reactions" subgroups). This hypothesis was not supported. The second hypothesis was that individuals receiving the high reaction instructions would show lower EPPS reliabilities when compared with individuals receiving the low reaction instructions. No support was found for this hypothesis. The third hypothesis that subjective reactions would be significantly related to state-anxiety was partially supported. The findings are discussed in terms of their implications for research in related areas as well as for further research on the relationship between subjective reactions and reliability. Reasons for the failure to achieve significant results are identified and possible ways of overcoming these difficulties are discussed. Areas for future research on this topic are identified.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION AND STATEMENT OF THE PROBLEM	1
II.	LITERATURE REVIEW	5
	A. Early Research	5
	B. Contemporary Research	7
	C. The Impact of Test Anxiety	10
	D. Characteristics of Test-Anxious Individuals	13
	E. The Hypotheses	22
III.	METHOD	24
IV.	RESULTS	28
V.	DISCUSSION	55
LIST OF REFERENCES		64
APPENDIXES		70
APPENDIX A.	REACTIONS QUESTIONNAIRE	71
APPENDIX B.	NUMBER OF PARTICIPANTS RESPONDING TO EACH CATEGORY OF TWELVE DEMOGRAPHIC TYPE ITEMS	81
APPENDIX C.	FREQUENCY AND PERCENTAGE OF PARTICIPANTS RESPONDING TO EACH CATEGORY OF EVERY ITEM OF THE EDWARDS PERSONAL PREFERENCE SCHEDULE	83
APPENDIX D.	ITEM MEANS, STANDARD DEVIATION, AND THE PERCENTAGE OF PARTICIPANTS RESPONDING TO EACH CATEGORY OF THIRTY-SIX REACTION QUESTIONNAIRE ITEMS	86

CHAPTER PAGE

APPENDIX E. PERCENTAGE OF PARTICIPANTS RESPONDING TO EACH OF FOUR CATEGORIES FOR EACH OF FORTY STATE-TRAIT ANXIETY INVENTORY ITEMS 88

APPENDIX F. LOADINGS OF THIRTY-SIX REACTION QUESTIONNAIRE ITEMS ON EACH OF FIVE REACTION FACTORS 89

VITA 90

LIST OF TABLES

TABLE PAGE

1. Percentage of Common Variance and Item Loadings on Each of Five Reaction Factors 30

2. Comparison of Obtained Means and Standard Deviations with Means and Standard Deviations from Normative Data. 37

3. Comparison of Means, Standard Deviations, and Internal Consistency Reliabilities (KR-20) for Dichotomized and Trichotomized Subgroups on Each of Five Reaction Factors for the Order Subscale of the Edwards Personal Preference Schedule 40

4. Comparison of Means, Standard Deviations, and Internal Consistency Reliabilities (KR-20) for Dichotomized and Trichotomized Subgroups on Each of Five Reaction Factors for the Affiliation Subscale of the Edwards Personal Preference Schedule 42

5. Comparison of Means, Standard Deviations, and Internal Consistency Reliabilities (KR-20) for Dichotomized and Trichotomized Subgroups on Each of Five Reaction Factors for the Endurance Subscale of the Edwards Personal Preference Schedule 44

6. Comparison of Means, Standard Deviations, and Internal Consistency Reliabilities (KR-20) for Dichotomized and Trichotomized Subgroups on Each of Five Reaction Factors for the Aggression Subscale of the Edwards Personal Preference Schedule 46

TABLE PAGE

7. Comparison of Means, Standard Deviations, and Internal Consistency Reliabilities (KR-20) on Two Instructional Sets for Each of Four Subscales of the Edwards Personal Preference Schedule 49

8. Comparison of Means and Standard Deviations for Two Instructional Sets for Each of Five Reaction Factors from the Reactions Questionnaire and the State-Anxiety and Trait-Anxiety Subscales of the State-Trait Anxiety Inventory 51

9. Intercorrelations of the Four Subscales of Edwards Personal Preference Schedule, Five Reaction Factors, and the Subscales of the State-Trait Anxiety Inventory. 53

10. Correlations of Selected Participant Characteristics with Five Reaction Factors, Four Subscales of the Edwards Personal Preference Schedule and the State-Anxiety and Trait-Anxiety Scales of the State-Trait Anxiety Inventory 54

CHAPTER I

INTRODUCTION AND STATEMENT OF THE PROBLEM

Psychological testing is widely used in our society, serving as an integral part of many personnel selection and assessment procedures in such varied areas as business, industry, education, and mental health. Goslin (1963) estimated that between 150 and 250 million tests are given each year in the United States. In part, as a result of the widespread use of testing and the potential for its misuse, public concern with testing increased during the 1960's. In 1965 Congressional Hearings were held to investigate psychological testing and its impact ("Testing and Public Policy," 1965). Much of the debate about testing at that time was directed at personality type instruments, with the primary concerns centered on issues such as invasion of privacy and the validity of the instruments (e.g., do they really measure what we think they do?).

The testing controversy is still with us today. However, the nature of the criticism has become more technically sophisticated and more narrowly focused over the years. For example, the use of testing in selection has come under increased scrutiny by various governmental agencies and has resulted in the development of a formal set of guidelines by which selection procedures will be evaluated ("Uniform Guidelines on Employee Selection Procedures," 1978).

With the increased emphasis on the technical standards of test construction and utilization, neither psychometricians nor test users

can afford to ignore situational or individual differences that result in decreased accuracy of measurement. However, little research in the past has dealt with how individual and situational characteristics interact to influence the psychometric properties of psychological measurement devices. Instead most efforts have either focused on improving test construction techniques and standardizing the conditions of administration or they have studied situational and individual characteristics that result in differential performance levels. Such efforts would seem to acknowledge the influence of subjective reactions on measurement; however, their influence on reliability has largely been ignored. This lack of research is especially significant when one considers the importance of reliability in measurement theory. Reliability is essential to measurement. With no reliability, we would have no confidence in what is measured. The well known standard error of measurement illustrates how the accuracy of a measure depends on its reliability. Thus, in the following equation

$$\sigma_E = \sigma_x \sqrt{1 - r_{xx'}}$$

σ_E represents the standard error of measurement for a measure $\underline{X}$; σ_X stands for the standard deviation of a measure; $\underline{r}_{xx'}$ represents the reliability coefficient for a measure. Therefore, it can be seen that as reliability decreases, the magnitude of the standard error of measurement increases. Also, in classical test theory the reliability of a test is a limiting condition for its validity. Consider, for example, the case where two fallible measures are correlated. The correction for unreliability is as follows:

$$\underline{r}_{xy} = \frac{\underline{r}'_{xy}}{\sqrt{\underline{r}_{xx'}}\ \sqrt{\underline{r}_{yy'}}}$$

where $\underline{r}_{xy}$ represents the "corrected" correlation between a variable x and a variable y; $\underline{r}_{xx'}$ represents the reliability of measure x; $\underline{r}_{yy'}$ represents the reliability of measure y; and $\underline{r}'_{xy}$ represents the observed correlation between the fallible x and y measures (Magnusson, 1967). Thus, the validity, $\underline{r}_{xy}$ is a direct function of the reliability of $\underline{X}$ and of $\underline{Y}$.

Stanley (1971) has listed several potential sources of variation in the testing situation which may contribute to lower accuracy (viz. lower reliability) of a psychological measure. These include: general characteristics of the individual (e.g., stable individual differences); "lasting but specific" individual characteristics (e.g., specific knowledge or past experience); temporary individual characteristics (e.g., health, fatigue, etc.); temporary specific individual characteristics (e.g., comprehension, memory, etc.); and systematic and chance factors (e.g., testing conditions) (p. 364). Subjective responses to each of these sources can be measured, thus it is possible to examine their influence on the reliability of a psychological measuring device.

The purpose of the present study was to investigate whether subjective reactions to a measurement device affect its reliability. The study can be conceptualized in three steps. The first step involves assessing subjective responses to the measurement device and the situation in which it is administered. Having measured individual responses with respect to the above sources of variation, it is next necessary to examine the underlying factor structure to reduce the dimensionality

of the items measuring reactions to measurement. Based on the resulting factors, the third step is to investigate the relationship between dimensions of subjective reaction to psychological measurement and the reliability of the measuring device. Here the emphasis is on determining whether reliability is lower when reactions are negative or unfavorable.

Although there is little research dealing directly with the present topic, there is an extensive body of research which is related to the general question of reaction to psychological measurement. This literature is reviewed in Chapter II.

CHAPTER II

LITERATURE REVIEW

A. Early Research

The earliest work that could be considered as dealing with reaction to measurement was conducted by physiologists during the first part of this century. The primary focus of these studies was the physiological manifestations of anxiety that resulted from being in an examination situation. For example; Folin, Demis, and Simillie (1914) studied medical students and found that approximately 18 percent of them passed sugar in their urine after a test, whereas there was no trace of sugar prior to the test. In a similar study, Cannon (1929) found that nearly half (44 percent) of a group of students taking a difficult examination passed sugar in their urine following the exam.

Luria (1932) noted speech and motor disturbances prior to examinations in students classified as unstable, while students classified as stable remained calm and did not show any unusual behavior patterns. Luria reported that nearly four times as many of the students classified as unstable were also classified as possessing neurotic symptoms. However, a major shortcoming of these early studies is that they did not control for individual differences and situational characteristics; that is, they simply took before and after measurements and did not consider intervening events; as a result, the findings are not easily interpretable.

Among the early studies that investigated individual reactions to testing were Benton (1935), Brown and Van Gelder (1938), Brown (1938A, 1938B), and Eisenberg (1941). Benton and Eisenberg were interested in how individuals vary in their interpretation of test items. Their findings indicated that interpretations varied as a result of individual differences and item ambiguity. For example, Benton found that approximately 31 percent of all individuals responding to the Personal Inquiry Form (a personality inventory) responded to individual items on the basis of a different interpretation of that item than did the majority of the respondents (analysis was conducted for each item individually). However, in comparing rates of differing interpretation between normal and abnormal respondents, Benton was not able to find a significant difference. Benton did find a significant negative relationship between the re-test reliability of an item(s) and scatter of interpretation of the item(s). As suggested by Friedenberg (1978), this finding would seem to offer support for the notion that the psychometric adequacy of a test can be influenced as a result of the way in which an individual responds to it.

In 1938 a series of articles (Brown and Van Gelder, 1938; Brown, 1938A; Brown, 1938B) reported the first use of a questionnaire to assess student reaction to a testing situation. The questionnaire contained seventy items that were designed to tap the subjective reactions of students to a test. Examples of the items include: "Do you get a sort of weak and helpless feeling before an examination?" "Are you irritable and hypersensitive before an examination?" (Brown, 1938A, p. 14). Brown's concern was that as classes became larger and

colleges and universities began to stress testing as the sole method of evaluating student performance there might be negative side effects resulting from the increased anxiety. Brown (1938B) reports a study in which he correlated results from the reactions questionnaire, neurotic tendency, intelligence, and grades on a comprehensive examination. He found a small negative correlation between intelligence and the reactions questionnaire (-.12) and between grades and the reactions questionnaire (-.22). The findings of this study suggest that individuals high in test-anxiety tended to do a little poorer on the test than individuals who were less anxious in the testing situation. (According to Brown the reactions questionnaire measured anxiety arising from the test/testing situation.) Brown's studies offer evidence supportive of many of the later findings which are reviewed below.

In addition to the studies previously cited, there was considerable interest in the concept of test-anxiety and reactivity to measurement in Europe during the 1930's. In Germany, for example, the first book on test-anxiety was published (Neuman, 1933), as well as a number of papers on the causes and treatment of test-anxiety (Redl, 1933; Weber, 1936). However, these studies never received much attention outside of Germany because they were not immediately translated (Spielberger, 1978).

B. Contemporary Research

The surge of research in the area of test-anxiety began in the United States around 1950 with a number of factors contributing to the rapid

emergence of research interest. As described by Spielberger (1978), some of these contributing factors were: The Meaning of Anxiety (May, 1950), Learning Theory and Personality Dynamics (Mowrer, 1950), and Personality and Psychotherapy (Dollard and Miller, 1950) and the development of the Manifest Anxiety Scale (MAS) by Taylor (1953). The MAS was the first widely used measure of the general trait of anxiety. However, because the MAS measured only the general trait of anxiety and was not sensitive to situational characteristics of anxiety it was not an appropriate measure for studying test-anxiety.

It was in part this shortcoming that lead Mandler and Sarason (1952, 1953; Sarason and Mandler, 1952) to develop the first widely used measure of test-anxiety. The Test Anxiety Questionnaire (TAQ) contained 39 items that attempted to measure the subjective experiences of anxiety that occur during testing. The results from the TAQ were correlated with grade point average (GPA), scores on college entrance exams (Scholastic Aptitude Test and Mathematical Aptitude Test), social class data (i.e., father's occupation, educational level, etc.), measures of frustration level (e.g., interfering responses), measures of intellectual ability, and behavioral manifestations of anxiety (i.e., perspiration, restlessness, etc.). The importance of Mandler and Sarason's work lies in the fact that they were able to demonstrate a measure of both concurrent validity and construct validity for the TAQ, thus, they were able to show reasonable support for the existence of test-anxiety. On the basis of their developmental work, Speilberger (1978) credits Mandler and Sarason with being the "pioneering researchers" in the area of test-anxiety research.

Following the development of the first standardized measure of test-anxiety, the TAQ (Mandler and Sarason, 1952, 1953; Sarason and Mandler, 1952), a number of other instruments have been developed to assess the presence and level of test anxiety. The Test Anxiety Scale (TAS) was developed by I. Sarason (1958) by using items from the TAQ and converting them to a true-false format. Both the TAQ and TAS are based on the assumption that test-anxiety is a learned behavior and that it interferes in problem-solving which in turn results in a decrement in performance for highly anxious participants.

Alpert and Haber (1960) developed the Achievement Anxiety Test (AAT). The AAT contains two scales; one scale measures the "debilitating" influence of anxiety and the other scale measures the "facilitating" influence of anxiety. The rationale for developing this scale was to allow for assessment of the differential impact of anxiety across individuals and situations.

The Worry-Emotionality Questionnaire (WEQ) developed by Liebert and Morris (1967) was designed to measure the worry and emotionality components of test anxiety. The worry component consists primarily of cognitive concerns about performance, while the emotionality items assess emotional reactions during testing. The WEQ is designed to be administered following a testing situation.

The State-Trait Anxiety Inventory (STAI) developed by Spielberger (1970) conceptualized anxiety as consisting of two principle components. The first component, trait anxiety, is a relatively long-term and enduring personality characteristic. The second component, state anxiety, is a temporary condition that fluctuates over time

and situations. State anxiety is intended as a measure of the situations in which test-anxiety is evoked.

The scales that have been discussed to this point have in common the conceptualization of test-anxiety as being a personality variable which is situationally evoked. In addition, the rationale behind the various scales conceived test-anxiety as interfering with task performance. The various measurement scales also have been constructed with one common additional characteristic: the potential influence of test-anxiety on the reliability of the testing instrument has been ignored.

C. The Impact of Test Anxiety

Perhaps the most important finding to emerge from the test-anxiety literature is the differential performance between high test-anxious and low test-anxious individuals. The differences in performance have been a consistent finding since the earliest studies (Brown, 1938B; Mandler and Sarason, 1952, 1953; Sarason and Mandler, 1952; I. Sarason, 1958; Sarason and Ganzer, 1962, 1963; Spielberger, 1972, 1978). Many variables have been examined to determine their relationship to level of anxiety and performance. In addition, several situational characteristics such as instructions, type of evaluation, attention to cues, etc., have been manipulated and the resulting influence on level of anxiety and performance studied. Each of these factors is discussed independently.

One explanation that has been made to account for the differential performance levels of high and low-anxious individuals is the manner in

which the situation is perceived. Findings from a number of studies have suggested that individuals characterized by high levels of test anxiety tend to perceive the testing situation as more personally threatening than do low test-anxious individuals (Mandler and Sarason, 1952; Sarason, 1958; Sarason, 1961; Wine, 1971). In these studies a number of variables have been investigated in an effort to identify those variables which account for the decrement in performance.

The influences of subjective success and subjective failure on anxiety level and test performance have been investigated in a number of studies (Mandler and Sarason, 1952; Doris and Sarason, 1955). In an effort to determine the impact of instructions on performance, Mandler and Sarason (1952) administered an anxiety questionnaire to 154 students. Following completion of the questionnaire, participants were told that they had performed very well, had done very poorly, or were told nothing. Performance was then measured on several trials of a second task. Results indicated that the high-anxiety group performed at a significantly lower level ($p < .05$) than did the low-anxiety group under conditions of both success and failure. These results indicate that a highly anxious participant's perception that his or her performance is being evaluated is sufficient to result in a decrement in performance. In addition, it may be that highly anxious individuals are more willing to accept and believe negative feedback.

In a similar study Doris and Sarason (1955) arbitrarily failed participants on several tests. After completing the tasks and receiving the arbitrary results, the participants were asked to rank order "blame statements." The blame statements included items

that directed blame for their performance to themselves or to others. When compared with scores from the TAQ it was found that high-anxiety participants attributed failure to themselves significantly more often ($p < .05$) than did the low-anxiety participants. These results would seem to indicate that highly anxious individuals tend to devalue their own performance. If actual differences in intellectual ability do not account for differences in performance as suggested by Brown (1938B) and Sarason and Mandler (1952), then it is likely that situational characteristics evoke behavior from participants that results in the decrement of performance. Such situational characteristics would also result in an increase in measurement error (e.g., true ability versus measured ability) and a decrease in the psychometric adequacy of the instrument.

In an effort to assess the impact of these subjective perceptions of self-worth, Sarason has conducted a series of studies using a verbal conditioning paradigm on college students who were extreme scorers on the TAS (Sarason and Ganzer, 1962, 1963; Sarason and Koenig, 1965). Participants were asked to describe themselves verbally during a thirty minute interview session with the researcher. During this session participants were under one of three conditions: "negative self-references" were reinforced, negative self-references were not reinforced, or "positive self-references" were reinforced. The results from these studies indicate that: (1) across conditions, the high test-anxious participants describe themselves more negatively than do participants who are less test-anxious, (2) when being reinforced for negative self-references the individual high in test-

anxiety is very responsive to reinforcement, (3) when individuals high in test-anxiety were reinforced for positive self-references they were unresponsive to reinforcement (Wine, 1971).

Mandler and Watson (1966) and Marlett and Watson (1968) assessed the extent to which individuals with extreme test-anxiety scores were bothered by perceptions of how they were doing while they were completing an assigned task. Results indicated that those individuals characterized by high test-anxiety scores were significantly more disturbed by thoughts of failure than were low test-anxious individuals.

D. Characteristics of Test-Anxious Individuals

The research findings reviewed above appear to indicate that highly test-anxious individuals are more likely to underrate their abilities, to be self-preoccupied and blame themselves for failures. One implication of these findings is that individuals who are plagued by feelings of self-doubt and fear of failure are forced to divide their attention between dealing with these feelings (task-irrelevant) and solving the task they are working on (task-relevant). One result of such a division of attention might be a decrement in performance. Another implication suggested by these findings is that high test-anxious individuals are more attentive to cues in the situations in which they are performing or being evaluated.

Test-anxiety has been compared with a number of performance variables to obtain a measure of the relationship between it and achievement and performance as measured by grade point average, Scholastic Aptitude Test, Mathematical Aptitude Test, intelligence and personality scales, and performance on other anxiety scales.

The goal of these comparisons, in part, has been to establish a measure of the validity for scales measuring the construct of test-anxiety. A clear understanding of the relationship among these variables is essential for interpreting the construct of test-anxiety.

In a study of 305 undergraduates Sarason (1957) found that test-anxiety was negatively correlated with grade point average (GPA) during the first two years of college ($p < .05$). Negative relationships were also found between test-anxiety and scores on the Scholastic Aptitude Test (SAT) ($p < .05$) and Mathematical Aptitude Test (MAT) ($p < .01$). A measure of general anxiety, the General Anxiety Scale (GAS) (Sarason, 1957) was found to be related to GPA during years one, two ($p < .01$), and four ($p < .05$) of college. These results are similar to those reported twenty years earlier by Brown (1938B). Evidence from these studies indicates that the correlation between test-anxiety and GPA is lower than the correlation between test-anxiety and scores on college entrance examinations (e.g., SAT, MAT). Sarason (1968) suggests that this may be an indication that students overcome certain aspects of being evaluated as a result of increasing familiarity with the components of the testing situations (e.g., teachers, requirements, etc.). These findings offer support to the notion that differences between high-anxious and low-anxious individuals are a result of situational characteristics rather than a result of true differences in the underlying abilities of individuals. Thus, by ignoring the influence of test anxiety, measures such as the SAT and MAT have increased error of measurement and lower reliability and validity.

Several investigations have approached test-anxiety as learned behavior (Mandler and Sarason, 1952, 1953; I. Sarason, 1960, 1972;

Berkeley and Sproule, 1973; Wine, 1971). Support for this notion will be cited from two sources. First, indications are that the impact of test-anxiety lessens with the passage of time. Sarason (1957), cited previously, noted that after the first two years of college, test-anxiety was no longer significantly related to GPA. Sarason, Pederson, and Nyman (1968) report a stronger relationship between test-anxiety and college entrance examination than between test-anxiety and GPA. This finding, as mentioned previously, indicates that being in an unfamiliar setting and being evaluated facilitates feelings of failing and as a consequence test-anxiety increases. As a result of the increase in test-anxiety attention is divided between relevant and irrelevant behaviors. With less attention focused on the task-relevant behaviors, performance is decremented.

One explanation for the decrement in the performance of high test-anxious individuals is that they may more actively seek cues from the environment to aid in problem-solving when they perceive themselves as being evaluated (Sarason, 1960; Sarason and Ganzer, 1963). As cited previously, studies of verbal behavior support the notion that high test-anxious individuals are more likely to be influenced by negative environmental cues or negative reinforcement and are less likely to respond to positive reinforcement than are less test-anxious individuals.

A second area of research relevant to the impact of learning and prior experience on test-anxiety comes from a number of studies that have examined the use of instructions to reduce the level of test-anxiety (Berkley and Sproule, 1973; Mandler and Sarason, 1953;

Phillips, 1972; Sarason, et al., 1960). In a study of 72 college sophomores who were extreme scorers on the TAQ, Mandler and Sarason (1953) varied the amount of prior experience. The experience conditions were: (1) no prior experience, (2) one previous trial, (3) two previous trials with the testing materials. Their findings indicate that as a result of increased experience, the variability of the high-anxious group decreased while their performance increased relative to the low-anxious group. The results seem to indicate that overcoming unfamiliarity with the setting and experiencing success enhances the performance of high-anxious participants.

In a review of test-anxiety literature, Berkley and Sproule (1973) concluded that individuals who are inexperienced test takers do not perform as well as individuals who are more sophisticated test takers. They summarize the literature by suggesting that the performance of high-anxious individuals on tests is likely not to reflect their true abilities. As suggested previously the high test-anxious individual may be forced to divide his/her attention between task-relevant and task-irrelevant cues and as a result performance is decremented. To overcome this problem, it was suggested that practice and/or coaching be used to increase the sophistication of inexperienced individuals in an effort to bring their performance in line with their true abilities. The end result of this would be to improve the validity of the test. Guion (1965) made a similar recommendation when he suggested that procedures be altered to allow high test-anxious individuals to obtain enough information about the test to reduce their level of anxiety. Guion suggested that this would help reduce the unreliability of the instrument.

The effects on performance of anxiety-arousing and anxiety-allaying instructions have also been studied. Mandelson (1973) administered a verbal learning task, Kunihihira's Japanese Word Learning Task, and a criterion test to a group of 96 seventh-grade students. To test the differential influence of instructions three experimental conditions were used: (1) "anxiety-arousing" instructions; (2) "anxiety-allaying" instructions; (3) control. Anxiety level, as measured by the Mood Adjective Check List (MACL), was found to be significantly higher for the anxiety-arousing group than for the anxiety-allaying group and the control group ($p < .01$). No significant difference was found between the anxiety level of the control group and the anxiety-allaying group. In addition, no significant overall differences were found in criterion test performance across the three experimental conditions. However, when performance was analyzed by sex of participant, significant differences did emerge. The anxiety scores for girls were found to be significantly higher under each of the experimental conditions ($p < .01$). Boys under the anxiety-arousing condition performed more poorly than boys under the allaying condition ($p < .01$). However, performance among the girls was found to be significantly higher under the anxiety-arousing condition ($p < .01$). Thus, it seems essential that anxiety level be broken down according to sex to determine whether significant differences do occur. In reviewing the literature, Phillips (1972) noted that lower class minority children were consistently found to have higher levels of test-anxiety than other lower class children. In addition, sex differences have been found to be more pronounced in lower class

and minority groups. These findings suggest that it is not sufficient to show differential performance among various levels of test-anxiety but that differences should be examined in respect to various demographic variables. Ignoring these factors and accounting for the differences as true differences in ability, run the risks of decreasing the reliability and validity of the instrument.

Other research efforts have examined the importance of the standardization of tests. For example, Fiske (1967) studied a simulated testing technique on a "nationally" representative sample of 589 individuals. Participants were given two of six possible simulated tests under one of two instructional conditions. The instructions asked participants to take the test as they would if they were applying for a job or taking part in research. No significant differences in reactions were found between the two conditions for the various tests. However, when asked to identify the purpose of the test, the modal response was generally correct. Based on the differential ability of individuals to identify the purpose of testing, Fiske concludes that a wide variety of reactions to testing should be expected. Fiske suggests that standardizing conditions of administration, while important, ignores variation resulting from individual reactions. Thus, it is important in the construction of tests to consider individual reactions. Anastasi (1967) suggested that part of the problem in test construction is that testing is becoming isolated from the mainstream of psychology. She suggested that psychometricians are devoting more time to "refining techniques of test construction, while losing sight of the behavior they set out to measure" (p. 297).

Tracy and Fiske (1974) examined the influence of additional information on the responses and attitudes of individuals taking a personality test. Participants were given varying amounts of information concerning a series of tests they were about to take. Their findings indicate additional information was associated with significantly more favorable attitudes toward the measuring instrument ($p < .001$). However, the results did not show a significant increase in response homogeneity or a reduction in carelessness of responding (e.g., proportion of missing responses) as information increased.

Friedenberg (1978), in reviewing the two previous studies (Fiske, 1967; Tracy and Fiske, 1974), suggested that the range of responses may have been restricted by the use of simulated (versus real) tests, given in an unrealistic setting, and by using a reactions questionnaire developed from free responses in a pilot study.

In summary, the most consistent finding to emerge from the test-anxiety literature is that individuals characterized by high and low levels of test-anxiety show different levels of performance. The decrement in performance of high test-anxious individuals has most adequately been accounted for by the suggestion that their attention is divided between task-relevant (the testing instrument) cues and task-irrelevant (environmental) cues. A number of research findings supports this notion. For example, as a group, high test-anxious individuals tend to perceive the testing/evaluation situation as more personally threatening and tend to blame themselves for failing. In addition, it has been found that highly anxious individuals tend to be more sensitive to, and more willing to believe, negative feedback.

Test-anxiety seems to be a learned behavior. For example, it has been shown that the influence of test-anxiety on performance can be lessened by providing the test-anxious individual additional information (experience) concerning the test/situation. Also, the impact of test-anxiety has been shown to decrease with the passage of time. Thus, just as test-anxiety can be acquired by learning (e.g., experiencing failure) it can be modified by learning (e.g., experiencing success).

In addition to the variables influencing test-anxiety and the characteristics of test-anxious individuals the literature was examined in an effort to identify situations in which test-anxiety could influence the reliability of a testing instrument. It was this latter topic that was the primary focus of this study. Two additional topics that were discussed in the literature review that were of interest in this study are the influence of instructions on performance and the relationship between certain participant characteristics (e.g., sex, age, GPA, etc.) and test-anxiety.

To overcome these difficulties, Friedenberg (1978) administered the General Aptitude Test Battery (GATB) to 180 men and women at a state employment office. Following administration of the GATB participants were asked to complete a reactions questionnaire. The reactions questionnaire was designed to identify factors that underlie subjective reactions to the GATB. It was hypothesized that higher reliabilities would emerge for individuals experiencing more positive reactions to the GATB and the GATB testing situation in comparison to those reporting negative reactions. The hypothesis was not supported, despite the fact that several distinct factors of reaction

to measurement did emerge. That reliability was unrelated to subjective reactions factors in Friedenberg's study may be attributable to the timed nature of the GATB, to its heavy emphasis on cognitive abilities (the measurement of which may be little affected by subjective reactions), and to the general tendency of persons to respond in a favorable way to the GATB and the surrounding situation.

In the present study an attempt was made to extend Friedenberg's research and to overcome some of the difficulties of his study by using a forced-choice personality instrument as the psychological measuring device to which subjects respond. It may be that negative reactions to measurement are more likely to affect the reliability of items measuring personality characteristics than items measuring cognitive abilities, since personality measurement tends to be somewhat influenced by situational factors (see, for example, Nunnally, 1978, Chapter 14). Also the nature of the forced-choice format reduces the ability of the test taker to respond toward items in an wholly positive manner. In addition, the procedure was untimed so all participants were allowed enough time to complete all items. As in the Friedenberg study, a questionnaire tapping reactions to measurement was administered immediately after administration of the measuring device. The Reactions Questionnaire used in this study was a modified version of the one developed by Friedenberg. With the exception of two additional items (questions 17 and 18) the item content of the Reaction Questionnaire on this study was the same as the one used previously by Friedenberg.

In addition, the present study attempted to manipulate subjective reactions to measurement directly through the instructions used for

administering the personality instrument to the subjects. Accordingly, two conditions were created, the first, the low reactivity condition, was written in an effort to minimize participant reaction by soliciting participation and describing the study in a polite and considerate manner. On the other hand, the second instructional set, the high reactivity condition, was worded in such a way as to evoke a more negative reaction from the participants. The rationale for using the instructional manipulation was to determine if the manner in which instructions were presented (e.g., tone of researcher, choice of words, etc.) had an impact on subsequent performance.

E. The Hypotheses

These hypotheses were developed in a manner consistent with the findings of the previous test-anxiety research such that the present study extends previous research efforts by examining the impact of subjective reactions to the testing instrument and testing situation. Based on the literature reviewed the following three hypotheses were examined in the present study.

H 1) When subjects are grouped on the basis of subjective reactions to measurement lower (higher) reliabilities for the personality instrument are observed for groups with less (more) favorable reactions.

H 2) When subjects are grouped on the basis of instructions, lower (higher) reliabilities for the personality instrument would be observed in the group receiving the high reactivity (low reactivity) instructions.

H 3) Subjective reactions to measurement are significantly related to state anxiety.

CHAPTER III

METHOD

Subjects

Questionnaires were administered to 237 undergraduates from the College of Business of the University of Tennessee. Questionnaires were received from 233 of the participants, who formed the sample for the present study. The participants consisted of 161 males and 72 females. All participants in this study were volunteers and were required to sign consent forms prior to participating. In addition, all university requirements for using human subjects were met and approval of the Human Subjects Committee was obtained prior to the start of the study.

Questionnaire

The questionnaire used in this study contained a total of 189 items in four major sections (see Appendix A). The first section contained twelve demographic type of items (e.g., sex of participant, age, grade point average, etc.). The second section contained 100 items forming four subscales of the Edwards Personal Preference Schedule EPPS). The four subscales were the Order, Affiliation, Endurance, and Aggression scales. The third section of the questionnaire consisted of the Reactions Questionnaire which contains 37 items in two sections which are designed to assess the subjective reactions of the participants to a psychological measurement device. The first

section contains 18 items which attempt to assess the impact of the conditions of administration (such as the room, instructions, noise level, people administering the scale, etc.). The second portion of the Reactions Questionnaire contains 18 items that attempt to assess reactions to the content of the EPPS items. The reactions questionnaire as mentioned previously was developed by Friedenberg (1978) in an effort to measure subjective reactions to measurement. To obtain a better measure of the sources of unreliability Friedenberg utilized the potential sources of variance listed by Stanley (1971, p. 364) in developing the reactions questionnaire.

The fourth section of the questionnaire contained the two 20-item scales from the State-Trait Anxiety Inventory (STAI) (Spielberger, 1970).

Procedure

The questionnaire was administered during regularly scheduled class meetings. In no instance was a class told that they would be filling out a questionnaire prior to the class meeting in which they actually completed the questionnaire.

For each of the seven classes the procedure was as follows. At the beginning of the class period the researcher introduced himself and told the class that they would have the opportunity to participate in a research study. Following this brief introduction the questionnaires were distributed and one of two sets of instructions was read aloud as the participants followed along on their questionnaires. After reading the instructions, the participants were instructed to proceed with the remainder of the questionnaire. The order of the

four sections of the questionnaire was held constant for all participants in the study. The order of the four sections was as follows: first, the demographic items; second, the EPPS items; third, the Reactions Questionnaire; and fourth, the STAI items. Questions concerning the initial instructions were answered. Questions concerning the purpose of the study or specific items in the questionnaire were deferred by asking the participant to answer the item as best they could or by asking them to save the question until everyone had completed the questionnaire. In addition, instructions were provided at the beginning of each of the last three sections of the questionnaire. These instructions provided specific directions for the portion of the questionnaire that followed and were designed to be self administering. After completing the questionnaire, participants were asked to check over their questionnaires to insure that they had not skipped a page or unintentionally left an item unanswered. The participants were then allowed to turn in their questionnaires and leave.

While completing the questionnaire, participants were free to leave any item unanswered if they felt that it was too personal or that it did not apply to them. In addition, participants were free to withdraw from the study at any time. Furthermore, it was stressed in the instructions that there would be no negative consequences for not participating or for leaving items unanswered. To insure that time was not a factor participants were given as much time as necessary to complete the questionnaire.

Instructional Manipulation

Two instructional conditions were used, low reactivity and high reactivity (see Appendix A for the exact text of the instructions in both conditions). The low reactivity condition was intended to solicit participation in a more polite and friendly manner. To do this, words and phrases such as "please" were used frequently and subjects were told "your participation is greatly appreciated." In the high reactivity condition instructions were intended to convey a feeling of rudeness and disrespect. To do this, the instructions were read in an unfriendly manner with no mention being made of any appreciation for their participation. In addition, an attempt was made to deindividualize responses by telling participants that we were not interested in who they were as individuals but were only concerned with their responses.

All members of a given class were given the same instructional set. The instructional set was varied between classes. The set of instructions a given class received was randomly determined prior to the start of the study. In all four classes (N = 136) received the high reactivity instructions while three classes (N = 97) received the low reactivity instructions.

CHAPTER IV

RESULTS

The frequency of responses for every item in each of the four sections of the questionnaire are presented in Appendices B-E.

Scale Scoring

Scale scores were developed for each of the subscales of the Edwards Personal Preference Schedule, the Reactions Questionnaire, and the State-Trait Anxiety Inventory. The procedures used to develop these scale scores are described below. In addition, a brief description of what each scale is intended to measure is provided.

The Edwards Personnel Preference Schedule

Scale scores were developed for each of the four EPPS subscales according to the procedures outlined in the EPPS manual (Edwards, 1959). For each subscale a correct response for a given item which was consistent with the subscale was scored "1" while an inconsistent response was scored "0". Scores for each subscale were then summed for each individual to form four composite scores--order, affiliation, endurance, and aggression.

A brief description of what each of the four EPPS scales measure is provided below. These descriptions are consistent with those provided in the EPPS manual (Edwards, 1959). For each of these four scales a higher score would be an indication that an individual

perceives the items that comprise the scale as more characteristic or descriptive of himself/herself than an individual who obtains a lower score. The order scale is intended as a measure of the degree of organization an individual experiences during day to day activities (e.g. planning the day's activities, doing certain activities at specified times, etc.). The affiliation scale was developed to assess the importance placed on friendship/attachments and the willingness of an individual to form new interpersonal relationships. Endurance is presented as a measure of an individual's willingness to finish whatever tasks he/she begins. The fourth scale, aggression, assesses aggressive tendencies by examining an individual's willingness to attack dissenting points of view, to respond with anger, and to blame others when things go wrong.

Reactions Questionnaire

To reduce the number of dimensions of the Reactions Questionnaire, a factor analysis was performed. The factor analysis procedure utilized for this phase of the analysis was principle factoring with iteration (PA2) from the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, and Bent, 1975). The PA2 method of factor analysis was selected because of its versatility and wide acceptance (Kim, 1975). The method of rotation selected to arrive at a final factor solution was the Varimax (orthogonal) procedure.

The factor analysis revealed five distinct reaction factors that accounted for 76.7 percent of the common variance. The factor composition and the percentage of common variance accounted for by each of these five factors after rotation are summarized in Table 1.

TABLE 1

PERCENTAGE OF COMMON VARIANCE AND ITEM LOADINGS ON EACH OF FIVE REACTION FACTORS

	Factors	Percentage of Common Variance	Factor Loadings
	1. General Evaluation of Scale Content	46.3	
Item 2	I wanted to take this scale.		.37
Item 19	The scale was interesting.		.50
Item 20	The scale allowed me to express my true feelings.		.50
Item 21	The scale appears to be a good measure of what it trys to measure.		.61
Item 23	On the whole, the scale was boring.		-.36
Item 31	The scale will provide information that will be put to good use.		.70
Item 33	My own answers will be important to the people who will analyze the results of the scale.		.73
Item 36	On the whole, I enjoyed filling out the scale.		.37
	2. Emotional Reaction to the Scale	11.0	
Item 16	The scale was given to me under false pretenses.		.48
Item 18	The instructions for the scale made me feel uncomfortable.		.63
Item 22	In general, the scale was an invasion of my privacy.		.53

TABLE 1 (Continued)

	Factors	Percentage of Common Variance	Factor Loadings
Item 24	The scale made me feel like a number rather than a human being.		.39
Item 30	The scale made me angry.		.39
Item 35	If I had it to do over again, I would change the way I answered a lot of the items.		.42
	3. Examinee Characteristics at the Time of Administration	7.5	
Item 8	The scale was given at a time when I was in a bad mood.		.48
Item 9	The scale was given at a time when I was in good health.		-.37
Item 10	The scale was given at a time when I was tired.		.77
Item 11	The scale was given at a time when I was mentally alert.		-.61
	4. Evaluation of Scale Characteristics	6.6	
Item 23	On the whole, the scale was boring.		.38
Item 25	The scale seemed silly and frivolous.		.44
Item 28	The scale was too long.		.70

TABLE 1 (Continued)

	Factors	Percentage of Common Variance	Factor Loadings
5.	Reaction to People Administering the Scale	5.5	
Item 14	The people administering the scale acted in an arrogant manner.		-.41
Item 15	The people administering the scale were friendly.		.88

Loadings of every item on each of the five factors are reported in Appendix F. To be included in the item composition of any factor it was determined that an item must have a loading of .35 or higher on that factor. Although the use of this cutoff point is arbitrary, it falls within the guidelines for determining the appropriate cutoff point suggested by Gorsuch (1974). The five factors that emerged are described below:

(1) Factor One--General Evaluation of Scale Content. With the exception of one item (item 2) the items loading on this factor are from the section of the questionnaire dealing with the content of the scale. In all, eight items loaded on this factor. The factor appears to reflect a "general evaluation of scale content." Examples of these evaluations are: "I wanted to take this scale," the scale was interesting and a good measure. Two additional items that reflect evaluations of scale content loaded on this factor. They were the importance of an individual's own responses, and the feeling that the information provided would be put to good use.

(2) Factor Two--Emotional Reaction to the Scale. The six items loading on this factor reflect emotional reactions to both the content of the scale and context of administration of the scale. Examples of the items loading on this scale are: "The instructions for this scale made me feel uncomfortable," and "The scale made me angry."

(3) Factor Three--Examinee Characteristics at the Time of Administration. The four items that load on this factor deal with personal characteristics of participants at the time the scale

was given. They are: "The scale was given at a time when I was in a bad mood," good health, tired, and mentally alert.

(4) Factor Four--Evaluation of Scale Characteristics. The three items that load on this factor come from the portion of the Reactions Questionnaire dealing with the content of the scale. The characteristics that emerged are: the scale was boring, too long, and seemed silly or frivolous.

(5) Factor Five--Reactions to the People Administering the Scale. Two items loaded on this factor. They are: "The people administering the scale acted in an arrogant manner," and "The people administering the scale were friendly." This factor appears to reflect subjective "reactions to the people administering the scale." However, the fact that the only two items loaded on this factor suggests that it should be interpreted with some caution.

Following the terminal factor solution, individual factor scores were derived from the factor-score coefficient matrix. The factor scores were computed for the purpose of dividing the total sample (N = 233) into subgroups. This foldback procedure which will involve splitting the rank ordered factor scores into dichotomous and trichotomous groups will be described in more detail below. However, it should be kept in mind that subsequent analyses will be based on subgroups of the same overall sample. Factor scores were computed according to the following procedure. First, the standardized values of the variables were computed [Z = (variable 1 - mean of variable 1) /

standard deviation of variable 1]. These standardized values were then multiplied by the "factor-score coefficient" for the same item and the resulting products were summed across all items to yield a composite factor score (Kim, 1975). This process was repeated for each individual. However, prior to computing the factor scores it was necessary to recode certain of the items on the Reactions Questionnaire to insure that a positive response to one item was equivalent to a positive response to all other items. For example, an item such as "The people administering this scale acted in an arrogant manner" was recoded such that to "disagree" with the item was the equivalent to responding "agree" to the statement, "The people administering the scale were friendly."

State-Trait Anxiety Inventory

Two scale scores were developed for the State-Trait Anxiety Inventory. The first, the State-Anxiety score was developed by summing the responses for the first twenty items for each individual. This scale was considered to be a measure of susceptibility to situationally induced anxiety such as might be encountered in a testing situation. The second scale, the Trait-Anxiety scale, was developed by summing responses for STAI items 21 to 40 for each individual Trait anxiety was considered to be a measure of the enduring personality trait of anxiety. Once again, negatively worded items were recoded such that a low score indicated that an individual possessed a lesser amount of the characteristic as measured by the scale. The recoded items are consistent with the reverse scored items identified in the STAI manual (Speilberger, 1970).

The means and standard deviations for scores on all of the subscales are presented in Table 2. In addition, the scale score means, standard deviations, and internal consistency reliabilities obtained in this study were compared with the means, standard deviations, and internal consistency reliabilities of the normative sample provided in the respective test manuals (Edwards, 1959, and Speilberger, 1970). The two sample means were compared by computing a *t* test. The assumption of equal group variances was tested by computing the ratio of the two variances (*F*-ratio). The internal consistency reliability coefficients were compared by using Fisher's *r* to *z* transformation. The comparisons described above are summarized in Table 2.

As can be seen from Table 2 the obtained scale means differed significantly from the scale means provided in the normative sample for three of the four EPPS subscales and the state anxiety scale of the STAI. In addition, the obtained reliabilities were significantly lower for each of the four EPPS scales.

Subscale Reliabilities

Using the full sample of respondents (N = 233), the internal consistency reliability for each of the subscales was interpreted by means of the KR-20 formula (Magnusson, 1967). The resulting KR-20 values for the EPPS were: Order--$\underline{r}_{xx}$ = .62, Affiliation--$\underline{r}_{xx}$ = .53, Endurance--$\underline{r}_{xx}$ = .75, Aggression--$\underline{r}_{xx}$ = .66. The two sections of the Reactions Questionnaire yielded the following internal consistency reliabilities (KR-20): the context section (items 1-18)--$\underline{r}_{xx}$ = .80 and the content section (items 19-36)--$\underline{r}_{xx}$ = .87. The KR-20

TABLE 2

COMPARISON OF OBTAINED MEANS AND STANDARD DEVIATIONS WITH MEANS AND STANDARD DEVIATIONS FROM NORMATIVE DATA

		N	Scale Mean	t-Value	Standard Deviation	F-Ratio	KR-20	Magnitude of Difference
ORDER	College Norms[1]	1509	10.24	2.90**	4.34	1.11	.74	3.19**
	Obtained Values	233	9.36***		4.12		.62	
AFFILIATION	College Norms[1]	1509	16.19	25.10**	4.36	1.19	.70	3.92**
	Obtained Values	233	8.57		4.00		.53	
ENDURANCE	College Norms[1]	1509	13.66	8.77**	5.14	1.32*	.84	3.51**
	Obtained Values	233	10.54		4.48		.75	
AGGRESSION	College Norms[1]	1509	11.70	-1.09	4.73	1.17	.84	6.05**
	Obtained Values	233	12.06		4.37		.66	
STATE ANXIETY	College Norms[2]	483	35.73	-1.90	9.46	1.16	.89	.57
	Obtained Values	233	37.13		8.77		.88	
TRAIT ANXIETY	College Norms[2]	483	37.96	-1.67	9.41	1.43*	.89	1.11
	Obtained Values	233	39.15		7.86		.87	

[1]Based on values provided in the manual for the Edwards Personal Preference Schedule (Edwards, 1959).

[2]Based on data provided in the manual for the State-Trait Anxiety Inventory (Spielberger, 1970).

* $p < .05$; ** $p < .01$

reliabilities of the two subscales of the STAI were: r_{xx} = .87 for the Trait-Anxiety scale and r_{xx} = .88 for the State-Anxiety scale.

Testing Hypothesis One

To test the first hypothesis, that different internal consistency reliabilities would be obtained as a result of grouping respondents according to subjective reactions, a comparison was made of the reliability of each of the four EPPS subscales between subgroups of participants based on the reaction factors. Two basic approaches were used to form subgroups--dichotomization and trichotomization. In the dichotomized groups, factor scores for each factor were rank-ordered, then the sample was split at the median. Splitting the sample in this manner resulted in two groups--a "high" group and a "low" group. The high group represented those participants who responded in a more favorable manner to the questionnaire/conditions of administration. This group was labeled the positive reactions group. The "low" group represented those individuals whose responses to the Reactions Questionnaire were less positive. This group was labeled the "negative reactions" group.

In developing the trichotomized groupings, a procedure similar to the procedure described above was utilized: the sample was divided into low, medium, and high subgroups on the basis of rank-ordered factor scores for each of the five reactions factors. However, for the purposes of this analysis comparisons were made only between "high" and "low" subgroups.

The means and standard deviations of the "high" and "low" groups were examined prior to the next phase of the analysis to insure

that neither group was influenced by differential restriction of range. This was accomplished by computing an F ratio between the high and low group variances for both the dichotomized and trichotomized splits. The resulting F values are presented in Tables 3-6. None of the F ratios exceeded a $\alpha = .05$ level of significance; therefore, it is concluded that any range restriction that might have occurred affected both groups equally.

During this phase of the analysis the means, standard deviations, and KR-20 reliabilities were compared for each subgroup according to each grouping procedure (trichotomized and dichotomized) for each of the five reactions factors. The analysis was repeated for each of the four EPPS subscales. The results are presented in Tables 3-6. Of the 40 comparisons the means differed significantly on only one occasion while none of the differences between standard deviations reached a $\alpha = .05$ level of significance.

The critical test for this hypothesis was the comparison of reliability coefficients between the "high-" and "low-reactions" groups for both the dichotomous and trichotomous splits. A total of twenty comparisons for each of the two grouping procedures were reported in Tables 3-6. The hypothesized relationship predicted that the low (negative reactions) group would show lower internal consistency reliabilities when compared to the high (positive reaction) group. However, in no instance was the magnitude of the difference between reliability coefficients statistically significant. Based upon the a priori prediction that the low-reaction group would show lower reliability a check was made of the direction of differences.

TABLE 3

COMPARISON OF MEANS, STANDARD DEVIATIONS, AND INTERNAL CONSISTENCY RELIABILITIES (KR-20) FOR DICHOTOMIZED AND TRICHOTOMIZED SUBGROUPS ON EACH OF FIVE REACTION FACTORS FOR THE ORDER SUBSCALE OF THE EDWARDS PERSONAL PREFERENCE SCHEDULE

Factor I		N	Group Mean	t-Value	Standard Deviation	F-Ratio	KR-20	Z
Dichotomized	High	116	9.23		4.16		.72	
				-.43		1.36		1.32
	Low	99	9.01		3.57		.62	
Trichotomized	High	62	8.57		3.93		.69	
				.54		1.27		0.83
	Low	58	8.91		3.48		.60	
Factor II								
Dichotomized	High	118	9.22		3.97		.70	
				-.37		1.06		0.40
	Low	88	9.03		3.85		.67	
Trichotomized	High	62	9.41		3.90		.68	
				-.89		1.03		0.10
	Low	60	8.82		3.84		.67	
Factor III								
Dichotomized	High	116	9.22		3.96		.69	
				-.36		1.05		0.13
	Low	90	9.03		3.86		.68	
Trichotomized	High	62	9.43		4.02		.70	
				.52		1.01		-0.11
	Low	62	9.07		4.05		.71	
Factor IV								
Dichotomized	High	114	9.18		4.16		.72	
				-.17		1.34		1.17
	Low	92	9.09		3.60		.63	
Trichotomized	High	60	8.97		3.77		.65	
				-.02		1.34		0.84
	Low	60	8.96		3.26		.55	

TABLE 3 (Continued)

		N	Group Mean	t-Value	Standard Deviation	F-Ratio	KR-20	Z
Factor V								
Dichotomized	High	113	9.24	-.46	4.00	1.10	.70	0.40
	Low	93	9.00		3.81		.67	
Trichotomized	High	59	9.15	-.56	3.64	1.00	.62	-0.09
	Low	63	9.22		3.65		.63	

TABLE 4

COMPARISON OF MEANS, STANDARD DEVIATIONS, AND INTERNAL CONSISTENCY RELIABILITIES (KR-20) FOR DICHOTOMIZED AND TRICHOTOMIZED SUBGROUPS ON EACH OF FIVE REACTION FACTORS FOR THE AFFILIATION SUBSCALE OF THE EDWARDS PERSONAL PREFERENCE SCHEDULE

		N	Group Means	t-Value	Standard Deviation	F-Ratio	KR-20	Z
Factor I								
Dichotomized	High	116	12.53		3.73		.58	
				-1.23		1.22		-0.90
	Low	87	11.85		4.11		.66	
Trichotomized	High	64	12.86		3.83		.60	
				-1.13		1.18		-0.63
	Low	58	12.04		4.16		.67	
Factor II								
Dichotomized	High	115	12.34		3.87		.61	
				0.35		1.05		-0.23
	Low	88	12.15		3.96		.63	
Trichotomized	High	60	11.82		3.97		.64	
				0.60		1.20		-0.58
	Low	59	12.28		4.36		.70	
Factor III								
Dichotomized	High	113	12.36		3.84		.60	
				-0.50		1.08		-0.45
	Low	90	12.08		3.98		.64	
Trichotomized	High	60	12.42		3.91		.62	
				-0.78		1.05		0.17
	Low	62	11.88		3.81		.60	
Factor IV								
Dichotomized	High	113	12.51		3.80		.59	
				-1.14		1.12		-0.80
	Low	90	11.88		4.03		.66	
Trichotomized	High	61	12.41		3.99		.62	
				-0.99		1.17		0.17
	Low	59	11.69		3.69		.60	

TABLE 4 (Continued)

		N	Group Means	t-Value	Standard Deviation	F-Ratio	KR-20	Z
Factor V								
Dichotomized	High	112	12.28	-0.18	4.00	1.12	.63	0.34
	Low	91	12.18		3.79		.60	
Trichotomized	High	60	12.53	-0.56	3.92	1.11	.62	0.33
	Low	60	12.13		3.71		.58	

TABLE 5

COMPARISON OF MEANS, STANDARD DEVIATIONS, AND INTERNAL CONSISTENCY RELIABILITIES (KR-20) FOR DICHOTOMIZED AND TRICHOTOMIZED SUBGROUPS ON EACH OF FIVE REACTION FACTORS FOR THE ENDURANCE SUBSCALE OF THE EDWARDS PERSONAL PREFERENCE SCHEDULE

		N	Group Means	t-Value	Standard Deviation	F-Ratio	KR-20	Z
Factor I								
Dichotomized	High	104	10.56		4.21		.71	
				0.98		1.29		-1.26
	Low	89	11.18		4.79		.79	
Trichotomized	High	57	9.61		4.08		.70	
				1.36		1.35		-0.60
	Low	57	10.72		4.74		.77	
Factor II								
Dichotomized	High	116	11.11		4.63		.77	
				-1.04		1.21		0.75
	Low	77	10.46		4.21		.72	
Trichotomized	High	63	11.40		4.98		.79	
				-1.20		1.35		0.94
	Low	49	10.36		4.28		.71	
Factor III								
Dichotomized	High	104	10.58		4.26		.73	
				0.92		1.22		-0.80
	Low	89	11.16		4.72		.78	
Trichotomized	High	53	10.66		4.49		.74	
				0.88		1.14		-0.49
	Low	60	11.43		4.80		.78	
Factor IV								
Dichotomized	High	110	11.37		4.57		.76	
				-1.96		1.17		0.46
	Low	83	10.14		4.22		.73	
Trichotomized	High	59	11.84		4.19		.68	
				-2.53*		1.08		-0.52
	Low	55	9.90		4.37		.73	

TABLE 5 (Continued)

		N	Group Means	t-Value	Standard Deviation	F-Ratio	KR-20	Z
Factor V								
Dichotomized	High	101	11.06		4.48		.75	
				-0.85		1.01		0.00
	Low	92	10.52		4.46		.75	
Trichotomized	High	54	10.84		4.62		.76	
				-0.17		1.14		0.46
	Low	61	10.70		4.32		.72	

TABLE 6

COMPARISON OF MEANS, STANDARD DEVIATIONS, AND INTERNAL CONSISTENCY RELIABILITIES (KR-20) FOR DICHOTOMIZED AND TRICHOTOMIZED SUBGROUPS ON EACH OF FIVE REACTION FACTORS FOR THE AGGRESSION SUBSCALE OF THE EDWARDS PERSONAL PREFERENCE SCHEDULE

		N	Group Means	t-Value	Standard Deviation	F-Ratio	KR-20	Z
Factor I								
Dichotomized	High	91	10.90		4.08		.66	
				0.12		1.02		-0.11
	Low	68	10.97		4.12		.67	
Trichotomized	High	52	11.10		3.92		.64	
				-0.32		1.02		0.00
	Low	41	10.85		3.96		.64	
Factor II								
Dichotomized	High	98	11.15		4.19		.69	
				-0.82		1.15		0.74
	Low	61	10.66		3.91		.62	
Trichotomized	High	53	11.51		4.23		.69	
				-0.32		1.10		0.41
	Low	43	11.27		4.04		.64	
Factor III								
Dichotomized	High	87	11.02		4.31		.70	
				-0.32		1.24		0.77
	Low	72	10.82		3.86		.63	
Trichotomized	High	48	11.76		3.88		.62	
				-1.46		1.02		-0.16
	Low	52	10.66		3.92		.64	
Factor IV								
Dichotomized	High	88	10.85		4.16		.68	
				0.29		1.06		0.22
	Low	71	11.03		4.05		.66	
Trichotomized	High	47	11.26		4.33		.70	
				-0.35		1.26		0.73
	Low	45	11.53		3.86		.61	

TABLE 6 (Continued)

		N	Group Means	t-Value	Standard Deviation	F-Ratio	KR-20	Z
Factor V								
Dichotomized	High	85	10.60		4.03		.65	
				1.26		1.06		-0.22
	Low	74	11.36		4.15		.67	
Trichotomized	High	44	10.85		3.70		.58	
				0.08		1.19		-0.62
	Low	54	10.79		4.04		.66	

The rationale for this comparison was that although the magnitude of the differences weren't significant support for the hypothesis could be obtained if the frequency of differences in the predicted direction exceeded chance expectations. This analysis indicated that for the dichotomous split 11 of 20 differences were in the predicted direction while 10 of 20 differences for the trichotomous split were in the predicted direction. Neither of these findings was statistically significant. Therefore, based upon the above analyses no support can be offered for the first hypothesis.

Testing Hypothesis Two

To test the second hypothesis, that different internal consistency reliabilities would be obtained as a result of grouping participants according to the type of instructional set they received, a comparison was made of the reliability of each of the four EPPS subscales between subgroups in the "high-" and "low-reactivity" instruction conditions. The means, standard deviations, and KR-20 reliability coefficients for each EPPS subscale were again computed in both instructional conditions. To test the assumption of equal variances an *F* ratio was computed using the "high-" and "low-reactivity" subgroup variances. Finally, the KR-20 reliability coefficients for both instructional groups on each of the four EPPS subscales were compared using Fisher's *r* to *z* transformation to check for different subgroup reliability to determine whether the reliability coefficients differed significantly and in the predicted direction. The results of the above analyses are summarized in Table 7.

TABLE 7

COMPARISON OF MEANS, STANDARD DEVIATIONS, AND INTERNAL CONSISTENCY RELIABILITIES (KR-20) ON TWO INSTRUCTIONAL SETS FOR EACH OF FOUR SUBSCALES OF THE EDWARDS PERSONAL PREFERENCE SCHEDULE

	N	Group Means	t-Value	Standard Deviation	F-Ratio	KR-20	z
ORDER							
Low Reactivity	86	9.01	-0.43	4.15	1.24	.72	0.92
High Reactivity	120	9.23		3.74		.65	
AFFILIATION							
Low Reactivity	84	11.85	-1.23	4.03	1.11	.65	0.57
High Reactivity	119	12.53		3.83		.60	
ENDURANCE							
Low Reactivity	80	11.18	0.98	4.16	1.30	.71	-0.58
High Reactivity	109	10.56		4.73		.76	
AGGRESSION							
Low Reactivity	63	10.97	0.12	4.02	1.07	.65	-0.32
High Reactivity	96	10.90		4.16		.68	

As can be seen from Table 7, no consistent differences emerged in the reliabilities between the "high-" and "low-reactivity" instruction groups. Thus, the high-reactivity instructional group did not show lower reliability as was predicted in hypothesis two. As a check on whether the instructional manipulation was effective in producing differential reactions to measurement, a comparison was made between the two instructional conditions using the mean scores and standard deviations for each of the five reaction factors and the state-anxiety and trait-anxiety subscales of the STAI. The comparison of scale means was achieved by the computation of t tests. In addition, a check was made for differential restriction of range between the two instructional conditions. The above analyses are summarized in Table 8.

The findings presented in Table 8 failed to show any significant differences in mean scores between instructional conditions. Thus, it appears that the instructional manipulation was not effective in producing different reactions to the personality instrument (e.g., lower reliability to the high reactivity instructions). In addition, no evidence of the existence of differential restriction of range was found. If restriction of range occurred it influenced both groups equally.

TABLE 8

COMPARISON ON MEANS AND STANDARD DEVIATIONS FOR TWO INSTRUCTIONAL SETS FOR EACH OF FIVE REACTION FACTORS FROM THE REACTIONS QUESTIONNAIRE AND THE STATE-ANXIETY AND TRAIT-ANXIETY SUBSCALES OF THE STATE-TRAIT ANXIETY INVENTORY

	N	Group Means	t-Value	Standard Deviation	F Ratio
Factor Score 1					
Low Reaction Condition	83	-0.0974	-1.29	0.890	1.01
High Reaction Condition	120	0.0674		0.895	
Factor Score 2					
Low Reaction Condition	83	0.0970	1.34	0.833	1.11
High Reaction Condition	120	0.0671		0.877	
Factor Score 3					
Low Reaction Condition	83	0.0628	0.87	0.850	1.02
High Reaction Condition	120	-0.0434		0.860	
Factor Score 4					
Low Reaction Condition	83	-0.0313	-0.45	0.793	1.13
High Reaction Condition	120	0.0216		0.841	
Factor Score 5					
Low Reaction Condition	83	0.0165	0.24	0.744	1.38
High Reaction Condition	120	-0.0114		0.875	
State-Anxiety					
Low Reaction Condition	96	37.52	0.91	9.42	1.37
High Reaction Condition	136	36.47		8.04	
Trait-Anxiety					
Low Reaction Condition	96	38.27	-1.00	7.70	1.13
High Reaction Condition	136	39.43		8.17	

Testing Hypothesis Three

To test the third hypothesis, that subjective reactions to measurement are significantly related to situation specific anxiety, Pearson correlation coefficients were computed by comparing the state-anxiety and trait-anxiety subscales of the State-Trait Anxiety Inventory with the factor scores computed for each of the five reaction factors for the entire sample (N = 233). These results are presented in Table 9. In addition, Table 9 contains the intercorrelations among the four EPPS subscales, the five reaction factors, and the two subscales of the STAI.

Of the results summarized in Table 9 it is of interest to note that all of the intercorrelations among the four EPPS exceeded a $\alpha = .05$ level of significance and that five of the six intercorrelations exceeded an $\alpha = .01$ level of significance. Also of interest is the fact that three of the five reactions factors correlated significantly with the state-anxiety subscale of the STAI. In addition, two of the five reactions factors correlated significantly with the trait-anxiety subscale of the STAI.

To facilitate interpretation of the relationship between the five reaction factors and participant characteristics Pearson correlations were computed. The result of this analysis is summarized in Table 10. It is of interest to note that of the various correlations reported in Table 10 that the variable sex correlated significantly on four of the five reactions factors and the affiliation and aggression subscales of the EPPS. In addition, participant grade point average was found to correlate significantly with each of the four EPPS subscales.

TABLE 9

INTERCORRELATIONS OF THE FOUR SUBSCALES OF EDWARDS PERSONAL PREFERENCE SCHEDULE, FIVE REACTION FACTORS, AND THE SUBSCALES OF THE STATE-TRAIT ANXIETY INVENTORY***

	Affiliation	Endurance	Aggression	Reaction Factor-1	Reaction Factor-2	Reaction Factor-3	Reaction Factor-4	Reaction Factor-5	State-Anxiety	Trait-Anxiety
Order	-.27**	.35**	-.20**	.01	.04	.01	.06	-.02	.02	.03
Affiliation		-.22**	-.25**	.08	-.04	.07	-.01	.01	.05	.06
Endurance			-.15*	-.03	.08	-.06	.22**	.01	.06	.06
Aggression				.06	.04	.03	-.00	.03	.02	.11*
Factor-1					.07	-.01	.06	-.05	.01	.08
Factor-2						-.07	.05	-.09	-.28**	-.23**
Factor-3							-.07	.05	.32**	.26**
Factor-4								-.07	-.07	.03
Factor-5									.16*	.06
State-Anxiety										.58**

* $p<.05$

** $p<.01$

*** as a result of the exclusion of missing data correlation coefficients were computed based on a sample that varied between N = 202 and N = 233

TABLE 10

CORRELATIONS OF SELECTED PARTICIPANT CHARACTERISTICS WITH FIVE REACTION FACTORS, FOUR SUBSCALES OF THE EDWARDS PERSONAL PREFERENCE SCHEDULE AND THE STATE-ANXIETY AND TRAIT-ANXIETY SCALES OF THE STATE-TRAIT ANXIETY INVENTORY***

	Sex	Age	Marital Status	Grade Point Average	Year in School
Reaction Factor -1	.21**	-.06	.07	.05	.02
Reaction Factor -2	.17*	.15*	.09	-.21**	.06
Reaction Factor -3	.12*	-.07	-.07	.10	.01
Reaction Factor -4	.09	-.08	.06	.01	-.06
Order	-.03	.20**	.20**	.11*	.13*
Affiliation	.25**	-.21**	-.17**	.15*	.03
Endurance	.03	.08	.02	.15*	-.00
Aggression	-.16**	-.00	-.00	-.13*	-.15*
State-Anxiety	-.09	-.04	-.07	-.12*	-.04
Trait-Anxiety	-.00	-.09	-.03	-.06	-.03

*$p < .05$

**$p < .01$

***as the result of the exclusion of missing data the sample size used in computing the correlation coefficients for this table varied between N = 227 and N = 233

CHAPTER V

DISCUSSION

The first hypothesis was that individuals who experienced negative (less positive) reactions to the testing instrument/ conditions of administration would show lower reliability on the subscales of the Edwards Personal Preference Schedule when compared to individuals experiencing more positive (less negative) reactions. Results of both the dichotomized split and the trichotomized split failed to show any overall pattern to support the notion that a relationship exists between high EPPS subscale reliabilities and positive reactivity and/or lower EPPS subscale reliabilities and negative reactivity. In addition, a comparison of the subgroup variances did not give an indication of any differential restriction of range.

The second hypothesis was that an experimental manipulation consisting of instructions designed to be highly reactive would have an adverse impact on the reliability of the instrument. The data failed to yield any consistent pattern supportive of the hypothesized relationship that lower (higher) reliabilities would be obtained in the high (low) reactivity instruction condition. One explanation for the absence of significant differences between the two instructional conditions is simply a reflection of the weakness of the high reaction instructions. Alternatively, it is possible that the lack of significant

differences was the result of the manner of presentation. For example, if the tone of the researchers voice did not convey the intended impression under the high reaction condition instructions then the expected effect of the instructions would be greatly lessened. However, regardless of the cause, if no differences existed between the treatments then no differences would be expected in the results.

The third hypothesis, that subjective reactions to measurement would be significantly related to situation specific anxiety, was tested by correlations between the two subscales of the STAI and the factor scores derived from the five reaction factors. Of the ten correlations reported in Table 9 (p. 53), five were statistically significant. Reaction factors two and three were significantly related to both the state- and trait-anxiety scales of the STAI. On the basis of these comparisons, the third hypothesis can be considered to be at least partially supported.

Although two of the main hypotheses of this study were not supported, a number of important findings did emerge. One important findings was the factor structure that emerged. The underlying factor structure is important for two reasons. First of all, the factors may be indicative of the existence of identifiable constructs dealing with reaction to measurement. The first factor which accounted for 46.3 percent of the variance and seems to deal with a general evaluation of the scale is consistent with Fishbein and Ajzen's (1975) contention that the first factor to emerge from a factor analysis of attitudes or beliefs is usually evaluative in nature. In addition, the general

nature of the first reaction factor from the present study is consistent with the first reaction factor to emerge in the previous study by Friedenberg (1978), although it should be noted that certain similarities in the factor structure would be expected between this study and the Friedenberg study because of the similarity of the Reactions Questionnaires that were used. The only difference in item content of the Reactions Questionnaire in the present study was the addition of items 17 and 18. An examination of reaction factors two, three, and four from the present study revealed negative impressions of the scale content and conditions of administration as well as negative reports of the participants' physical conditions at the time of administration. These findings are significant in that one of the goals of this study was to reduce the tendency to respond to the testing instrument in a solely positive manner by using the forced-choice items from the EPPS. This finding is further clarified by comparing the overall percentage of negative responses (see Appendix D, p. 86) to the Reactions Questionnaire in this study (25 percent) with the percentage of negative responses previously reported by Friedenberg (13 percent) ($p < .05$).

The reaction factors identified in this study could be used in a variety of settings employing psychological measurement as well as to form the basis for further research. For example, in an applied setting, one potential use of the Reactions Questionnaire would be to identify potentially distracting conditions in conjunction with a selection procedure (e.g., the presence of other people in the room, physical aspects of the room, mannerisms of an interviewer, etc.).

In analyzing the study a number of possible explanations can be offered for the absence of significant results. While it is apparent that no single explanation would be sufficient to explain the failure to achieve significant results, it is important to examine the possible impact that certain of the procedures utilized in this study had on the final outcome.

The participants selected for this study were undergraduates enrolled in courses in the college of business. While having such a "captive" subject pool facilitated data collection, it may not have been without certain costs. For instance, there is no way to insure that the sample used in this study is representative of the total population. It is possible that when compared with the larger population, the sample of students would be skewed toward one end of the distribution. As a group the students may be more homogeneous and therefore the results could have been influenced by restriction of range. The greater homogenity of the present sample as compared to the general population is evidenced by the comparison of the obtained results with the normative data summarized in Table 2 (p. 37). As was reported, three of the four EPPS subscale means were significantly lower in the present sample as compared to the college norms provided in the EPPS manual (Edwards, 1959). Also, each of the four EPPS subscale reliabilities were significantly lower in this study as compared with the normative data for college students provided in the manual. The magnitude of the differences between the obtained scale means and the scale means of the normative sample may in part reflect changes in certain characteristics of the general population between the time the normative sample was made and the present time.

Another factor that could have adversely influenced the results is the amount of experience college students have had in test taking situations. As a result of their increased amount of experience with testing, college students may be desensitized to testing since they have participated in a wide variety of testing situations. If this were true, it seems that the conditions of administration would be less of a factor for college students in the present study because participants were given the questionnaire in a familiar setting (e.g., a class in which they have been enrolled for approximately four weeks).

In addition to their increased experience with testing, college students in many cases have had considerable contact with individuals conducting research. One response on the part of students to their frequent use as subjects is a feeling of resentment. Another response, which perhaps is even more common is not answering the questionnaire seriously. Considering the conditions of administration there is little that could have been done to prevent such response characteristics. In retrospect the instrument was presented as a research instrument in which there were no consequences, positive or negative, for completing the questionnaire, therefore, the questionnaire lacked certain elements of realism.

It is, of course, possible that the instrument (EPPS) used in this study may not have been a good choice for trying to measure differences in reliability. The question of how reactivity should relate to performance on the EPPS is unclear. That is, while an individual may react negatively to the instrument and/or conditions of administration, this does not mean that corresponding attenuation or

differences will occur for EPPS scores. Also, in a personality test such as the EPPS specific knowledge is not required, rather the individual is in a position to provide self-report type of information. The individual is in a position to present him/herself in a favorable manner. Therefore, the differences between individuals who experience negative and positive reactions to the EPPS could potentially be confounded. In addition, only four subscales were used for the purposes of this study so even if the EPPS were an adequate personality measure and there were measureable personality differences between high and low reactivity individuals using the abbreviated form (four scales) would not provide a complete picture of an individual's personality.

Implications

While the results of this study did not support the hypothesized relationships a number of implications for future research did emerge. One area that needs to be explored more fully is the underlying constructs that account for the reaction factors that were identified in the factor analysis. The findings of both the present study and the earlier study by Friedenberg (1978) suggest that a clear and interpretable factor structure does exist. However, much additional research is needed in order to develop explanations and to identify the factors that influence subjective reactions to testing and/or the conditions of administration. The goal of such research would, in part, be to identify those situations and/or test characteristics under which an individual's performance deviates from his/her "true" ability, resulting in decreased reliability of measurement.

One potentially encouraging implication of this study is the fact that (personality) tests may be more robust than they are sometimes given credit for in terms of the impact that negative participant reactions/characteristics have on performance. For instance, the results presented from this study indicate that those individuals that reported negative reactions to the instrument/conditions of administration did not differ significantly in their performance from the group with more positive reactions (e.g., did not result in lower mean scale score or reliability). The implications of this finding would suggest that additional research should be conducted to determine how critical the emphasis on standardization is.

An alternative explanation for the lack of significant differences between instructional conditions could be offered on the basis of the previously discussed failure of the instructional conditions to result in different levels of reaction. This failure was attributed to the weakness of the manipulation. However, an alternative explanation could be offered that the absence of significant differences was the result of the wide range of instructional conditions that could have been presented without adversely influencing participant responses. This alternative explanation suggests that the emphasis on standardization needs to be examined to determine under what conditions and with which type of tests (e.g., personality, objective, etc.) is standardization most critical. For example, how critical are considerations such as lighting in the testing room, friendliness of the individual administering the test, and the amount of detail provided in the instructions

in "real world" settings? Research of the type described above would be consistent with Anastasi's (1967) suggestion that steps be undertaken to bring the work of psychometricians more into the mainstream of psychology for the purposes of test development.

One problem area that needs to be overcome in future studies is to use the Reactions Questionnaire in conjunction with an instrument that is more indicative of performance. In addition, it seems as if the more realistic the situation the more likely significant (e.g., interpretable) results will be obtained. Ideally such a situation might involve some potentially negative consequences for poor performance. For example, such a situation might be provided by the consequences of performing poorly on a selection battery or a standardized test such as the Graduate Record Examination where the individual does not get a certain job or into graduate school as a result of poor performance. One of the principal benefits of administering the Reactions Questionnaire in the type of situation described above would be the improved generalizability of results. Thus, future research with the Reactions Questionnaire should focus on measuring reactions to a more realistic instrument in a more realistic situation (e.g., with an actual test).

The exclusive use of college students should be avoided in favor of using a less homogeneous sample more representative of the total population. However, this is not intended to mean students should be ignored; rather it is intended that the focus be enlarged to include some non-academic testing situations as a part of a larger cross-sectional study.

One possibility for a future study using college students would be to administer the Reactions Questionnaire following in class tests. The use of the Reactions Questionnaire as suggested above would allow for a large data base. In addition, because the Reactions Questionnaire can be completed quickly the data base could include reactions to various types of tests (i.e., essay, multiple-choice, etc.). Comparisons could then be made between the various types of tests as well as between year in school (e.g., freshman, sophomore, junior, senior, etc.) to determine whether differences do exist in reactions to measurement. However, as suggested previously it would be desirable to collect data in non-academic settings. This would make possible comparisons to determine if there are differences in the underlying reactions to measure of students and non-students.

Finally, in future research efforts it might be worthwhile to more carefully assess the relationship of variables such as sex of participant, grade point average, and socio-economic status with level of test-anxiety to determine if and how they interact to influence performance. Such findings would be of interest in interperting the results of standardized tests that show different performance levels for males and females or different racial and socio-economic groups. Research of this type would do much to identify the underlying factors that influence reactions to testing.

LIST OF REFERENCES

LIST OF REFERENCES

Alpert, R., and Haber, R. N. Anxiety in academic achievement situations. Journal of Abnormal and Social Psychology, 1960, 61, 207-215.

Anastasi, A. Psychology, psychologists, and psychological testing. American Psychologist, 1967, 22 (4), 297-306.

Benton, A. L. The interpretation of questionnaire items in a personality schedule. Archives of Psychology, 1935 (May), 190.

Berkley, C. S., and Sproule, C. F. Test-anxiety and test unsophistication: The effects, the cures. Public Personnel Management, 1973, January-February, 55-59.

Brown, C. H. Emotional reactions before examinations: The results of a questionnaire. Journal of Psychology, 1938, 5, 11-26.

Brown, C. H. Emotional reactions before examinations: Intercorrelations. Journal of Psychology, 1938, 5, 27-31.

Brown, C. H., and Van Gelder, D. Emotional reactions before examinations: Physiological changes. Journal of Psychology, 1938, 5, 1-9.

Cannon, W. B. Bodily changes in pain, hunger, fear, and rage. New York: Appleton, 1929.

Dollard, J., and Miller, N. E. Personality and psychotherapy. New York: McGraw-Hill, 1950.

Doris, J., and Sarason, S. B. Test anxiety and blame assignment in a failure situation. Journal of Abnormal and Social Psychology, 1955, 50, 335-338.

Eisenberg, P. Individual interpretation of psychoneurotic inventory items. Journal of General Psychology, 1941, 25, 19-40.

Edwards, A. L. Edwards personal preference schedule manual. New York: The Psychological Corporation, 1959.

Fishbein, M., and Ajzen, I. Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, Pa.: Addison-Wesley, 1975.

Fiske, D. W. The subject reacts to tests. American Psychologist, 1967, 22 (4), 287-296.

Friedenberg, I. J. Reaction to psychological measurement. Unpublished Master's thesis, University of Tennessee, 1978.

Goslin, D. A. The search for ability: Standardized testing in social perspective. New York: Russell Sage Foundation, 1963.

Gorsuch, R. L. Factor analysis. Philadelphia: W. B. Saunders, 1974.

Guion, R. M. Personnel testing. New York: McGraw-Hill, 1965.

Haggard, E. A. Intraclass correlation and the analysis of variance. New York: The Dryden Press, Inc., 1958.

Hase, H., and Goldberg, L. Comparative validity of different strategies of constructing personality inventory scales. Psychological Bulletin, 1967, 67 (4), 231-248.

Kim, J. Factor analysis. In Nie, N. H., Hull, C. H., Jenkins, J., Steinbrenner, K., and Bent, D. H. (Eds.), Statistical package for the social sciences. New York: McGraw-Hill, 1975.

Kleinbaum, D., and Kupper, L. Applied regression analysis and other multivariable methods. North Scituate, Massachusetts: Duxbury Press, 1978.

Liebert, R., and Morris, L. Cognitive and emotional components of test anxiety: A distinction and some initial data. Psychological Reports, 1967, 20, 975-978.

Luria, A. R. The nature of human conflicts. (W. H. Gantt, Trans.) New York: Liveright, 1932.

Magnusson, D. Test theory. (H. Mabon, Trans.) Reading, Mass.: Addison-Wesley, 1966.

Mandelson, L. R. Test performance on a verbal learning task as a function of anxiety-arousing testing instructions. Journal of Educational Research, 1973, 67 (1), 37-40.

Mandler, G., and Sarason, S. B. A study of anxiety and learning. Journal of Abnormal and Social Psychology, 1952, 47, 166-173.

Mandler, G., and Watson, D. L. Anxiety and the interruption of behavior. In C. D. Speilberger (Ed.), Anxiety and behavior. New York: Academic Press, 1966.

Marlett, N. J., and Watson, D. Test anxiety and immediate or delayed feedback in a test-like avoidance task. Journal of Personality and Social Psychology, 1968, 8, 200-203.

May, R. The meaning of anxiety. New York: Ronald Press, 1950.

Megargee, E. I. The California psychological inventory handbook. San Francisco: Jossey-Bass, Inc., 1972.

Mowrer, O. H. Learning theory and personality dynamics. New York: Ronald Press, 1950.

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

Neumann, J. Anxiety and illness before examination. Berlin: Gutersloh, 1933.

Nunnally, J. C. Psychometric theory. New York: McGraw-Hill, 1967.

Nunnally, J. C. Psychometric theory (2nd ed.). New York: McGraw-Hill, 1978.

Redl, F. We teachers and the fear of examinations. Z. Psychoanal. Padag., 1933, 7, 398-400.

Sarason, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. Journal of Experimental Psychology, 1956, 51, 253-260.

Sarason, I. G. Test anxiety, general anxiety, and intellectual performance. Journal of Consulting Psychology, 1957, 21 (6), 485-490.

Sarason, I. G. The effects of anxiety, reassurance, and meaningfulness of material. Journal of Experimental Psychology, 1958, 56 (6), 485-490.

Sarason, I. G. A note on anxiety, instructions, and word association performance. Journal of Abnormal and Social Psychology, 1961, 62, 153-154.

Sarason, I. G. Verbal learning, modeling, and juvenile delinquency. American Psychologist, 1968, 4, 254-266.

Sarason, I. G. Experimental approaches to anxiety: Attention and the uses of information. In C. D. Spielberger (Ed.), Anxiety: Current trends in theory and research (Vol. 2). New York: Academic Press, 1972.

Sarason, I. G., and Ganzer, V. J. Anxiety, reinforcement, and experimental instructions in a free verbal situation. Journal of Abnormal and Social Psychology, 1963, 67, 87-91.

Sarason, I. G., and Koenig, K. P. The relationship of test anxiety and hostility to description of self and parents. Journal of Personality and Social Psychology, 1965, 2, 617-621.

Sarason, I. G., and Minard, J. Interrelationships among subject, experimenter, and situational variables. Journal of Abnormal and Social Psychology, 1963, 67, 87-91.

Sarason, I. G., and Palola, E. G. The relationship of test and general anxiety, difficulty of task, and experimental instructions to performance. Journal of Experimental Psychology, 1960, 59 (3), 185-191.

Sarason, I. G., Pederson, A. M., and Nyman, B. Test anxiety and the observation of models. Journal of Personality, 1968, 36, 493-511.

Sarason, S. B., and Mandler, G. Some correlations of test anxiety. Journal of Abnormal and Social Psychology, 1952, 47, 810-817.

Sieber, J. E. Effects of decision importance on ability to generate warranted subjective uncertainty. Journal of Personality and Social Psychology, 1974, 30 (5), 688-694.

Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. STAI manual. Palo Alto: Consulting Psychologists Press, 1970.

Spielberger, C. D. (Ed.). Anxiety: Current trends in theory and research. (Vol. I). New York: Academic Press, 1972.

Spielberger, C. D. (Ed.). Anxiety: Current trends in theory and research. (Vol. II). New York: Academic Press, 1972.

Spielberger, C. D., and Sarason, I. G. (Eds.). Stress and anxiety. (Vol. 5). New York: Wiley, 1978.

Stanley, J. C. Reliability. In R. L. Thorndike (Ed.), Educational measurement. Washington, D.C.: American Council of Education, 1971.

Taylor, J. A. A personality scale of manifest anxiety. Journal of Abnormal and Social Psychology, 1953, 48, 285-290.

Testing and public policy. (Special issue). American Psychologist, 1965, 20, 857-993.

Tracy, R. L., and Fiske, D. W. Does information about personality scales affect responses and attitudes? Educational and Psychological Measurement, 1974, 34, 617-629.

Uniform guidelines on employee selection procedures. Federal Register, Aug., 1978.

Weber, C. Examination anxiety in taking information tests. Industrial Psychotechnology, 1934, 11, 53-55.

Wine, J. Test anxiety and direction of attention. Psychological Bulletin, 1971, 76 (2), 92-104.

APPENDIXES

APPENDIX A

REACTIONS QUESTIONNAIRE

Instruction:

Low Reactivity Condition

I am doing a study for my masters thesis and I would really appreciate it if you would help me by filling out a questionnaire. I am attempting to develop a measuring instrument that will allow me to study the relationship among various personal characteristics. To do this, I really need your help in filling out the questionnaire. This questionnaire is being developed strictly for research purposes.

The research deals with the role of individual differences in relationships among various personal attributes and personality dimensions.

I would like to refine the questionnaire to the point that it would allow me to accurately identify different personality characteristics. For this reason, I need your cooperation in filling out questionnaire by answering each question carefully and honestly. All that is involved in filling out the questionnaire is indicating your response to each of a series of questions. You may leave any question unanswered if you feel that it is too personal or if you feel that it doesn't apply to you. In addition, you are free to withdraw from the study at anytime. All I ask is that you please read each item carefully. Usually your first response is an accurate one, so don't spend too long on any item.

If you are concerned about the confidentiality of your responses, I assure you no effort will be made to identify individual responses. I guarantee that all of your responses will be kept in strictest confi-

dence. If you would like to participate, please read the consent form on the next page and sign it. The consent form will be removed from the questionnaire to insure the anonymity of your responses. Once again, thank you for your cooperation. I really appreciate it.

Are there any questions before we begin?

Instruction:

High Reactivity Condition

I am doing a study for my masters thesis and I would like you to help me by filling out a questionnaire. I am attempting to develop a measuring instrument that can identify latent personality characteristics. The personality characteristics that we are especially interested in identifying are those that identify individuals who may not adjust well to college life. Although this instrument is being developed for research purposes, it might be beneficial to have a device which could be used to identify people who don't fit into college life.

I would like to refine the questionnaire to the point that it could be used by instructors to identify those individuals in their classes who may be reacting adversely to various situations. You can help me by taking the time to fill out a questionnaire. All that is involved in filling out the questionnaire is indicating your response to each of a series of questions. You may leave any question unanswered if you feel that it is too personal or if you feel that it doesn't apply to you. In addition, you are free to withdraw from the study at anytime. All I ask is that you not take all day filling out these forms. I don't have time to waste so please hurry.

If you are concerned about confidentiality, don't worry. I am not concerned with who you are as individuals. All I am interested in is your answers. I guarantee that all of your responses will be kept in strictest confidence. If you would like to participate,

please read the consent form on the next page and sign it. The consent form will be removed from the questionnaire to insure the anonymity of your responses.

INSTRUCTIONS: Now we would like to ask you some questions about what you thought about the section of the scale you just filled out (hereafter referred to as the scale) and how you felt while taking it. The information you provide will help us evaluate the usefulness of the scale and how well it was administered.

Please read the statements below carefully and indicate how much you agree or disagree with each one of them. Agreement and disagreement can be shown by circling one of the seven numbers following the statement. The numbers range on the scale from 1 to 7 as follows:

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Tend to Disagree	Uncertain	Tend to Agree	Agree	Strongly Agree

Thus, if you strongly agree with a statement, circle the number 7 to the right of the statement. Or, as another example, of you tend to disagree with a statement, circle the number 3.

(Please circle one number for each statement)	STRONGLY DISAGREE	DISAGREE	TEND TO DISAGREE	UNCERTAIN	TEND TO AGREE	AGREE	STRONGLY AGREE
A. HOW AND WHEN THE SCALE WAS GIVEN							
1. The purpose of the scale was adequately explained to me	1	2	3	4	5	6	7
2. I wanted to take this scale	1	2	3	4	5	6	7
3. I was allowed sufficient time to think through my answers	1	2	3	4	5	6	7
4. There were too many people in the room when I filled out the scale	1	2	3	4	5	6	7
5. It was easy to be distracted while filling out this scale	1	2	3	4	5	6	7
6. The scale was given in a physically comfortable setting	1	2	3	4	5	6	7
7. The scale was given at a convenient time and location	1	2	3	4	5	6	7
8. The scale was given at a time when I was in a bad mood	1	2	3	4	5	6	7
9. The scale was given at a time when I was in good health	1	2	3	4	5	6	7
10. The scale was given at a time when I was tired	1	2	3	4	5	6	7
11. The scale was given at a time when I was mentally alert	1	2	3	4	5	6	7
12. I really had no choice as to whether or not I would take this scale	1	2	3	4	5	6	7
13. I was not at all nervous when filling out the scale	1	2	3	4	5	6	7

(Please circle one number for each statement)	STRONGLY DISAGREE	DISAGREE	TEND TO DISAGREE	UNCERTAIN	TEND TO AGREE	AGREE	STRONGLY AGREE
14. The people administering the scale acted in an arrogant manner.....	1	2	3	4	5	6	7
15. The people administering the scale were friendly.....	1	2	3	4	5	6	7
16. The scale was given to me under false pretenses.....	1	2	3	4	5	6	7
17. The instructions for the scale were clear and understandable to me.....	1	2	3	4	5	6	7
18. The instructions for the scale made me feel uncomfortable.....	1	2	3	4	5	6	7
B. THE CONTENT OF THE SCALE							
19. The scale was interesting.....	1	2	3	4	5	6	7
20. The scale allowed me to express my true feelings.....	1	2	3	4	5	6	7
21. The scale appears to be a good measure of what it trys to measure.....	1	2	3	4	5	6	7
22. In general, the scale was an invasion of my privacy.....	1	2	3	4	5	6	7
23. On the whole, the scale was boring.....	1	2	3	4	5	6	7
24. The scale made me feel like a number rather than a human being.....	1	2	3	4	5	6	7
25. The scale was frivolous.....	1	2	3	4	5	6	7
26. I really had to pay attention when completing this scale.....	1	2	3	4	5	6	7

(Please circle one number for each statement)	STRONGLY DISAGREE	DISAGREE	TEND TO DISAGREE	UNCERTAIN	TEND TO AGREE	AGREE	STRONGLY AGREE
27. The scale cannot measure my unique characteristics.....	1	2	3	4	5	6	7
28. The scale was too long.....	1	2	3	4	5	6	7
29. Some of the questions made me feel foolish.....	1	2	3	4	5	6	7
30. The scale made me angry.....	1	2	3	4	5	6	7
31. The scale will provide information that will be put to good use.....	1	2	3	4	5	6	7
32. Most of the items in the scale were clear and understandable.....	1	2	3	4	5	6	7
33. My own answers will be important to the people who will analyze the results of the scale.....	1	2	3	4	5	6	7
34. I would be willing to fill out the scale again.....	1	2	3	4	5	6	7
35. If I had to do over again, I would change the way I answered a lot of the items.....	1	2	3	4	5	6	7
36. On the whole, I enjoyed filling out the scale.....	1	2	3	4	5	6	7

37. Other comments about the scale: _____

_____

_____

BACKGROUND INFORMATION

I. Your sex: 1. Male 2. Female (circle one)

II. Your age:
1. 18 or under
2. 19-21
3. 22-24
4. 25-30
5. 31-40
6. 41 or over

III. Marital status:
1. single
2. married
3. separated, divorced, or widowed

IV. What is your cumulative grade point average (G.P.A.) at the present time? (circle one)

1. less than 2.0
2. 2.00 - 2.50
3. 2.51 - 3.00
4. 3.01 - 3.50
5. 3.51 - 4.00
6. Don't know

V. Approximately, what do you think your final (cumulative) grade point average (G.P.A.) will be when you graduate from U.T.? (circle one)

1. less than 2.0
2. 2.00 - 2.50
3. 2.51 - 3.00
4. 3.01 - 3.50
5. 3.51 - 4.00
6. Don't know

VI. What final grade do you expect to receive in this course? (circle one)

1. A
2. B
3. C
4. D
5. F
6. Audit/No credit

VII. Approximately how many friends do you have that you associate with on a regular basis? (circle one)

1. None
2. 1-2
3. 3-5
4. 6-8
5. 9-12
6. 13 or more

VIII. How many clubs or organizations are you a member of?

1. None
2. One
3. Two
4. Three
5. Four
6. Five or more

IX. How many elected offices have you held in the various clubs and organizations of which you have been a member?

1. None
2. One
3. Two
4. Three
5. Four
6. Five or more

X. What is your current status in school? (circle one)

1. Full-time 2. Part-time

XI. What year of college are you in?

1. Freshman 3. Junior 5. Graduate student
2. Sophomore 4. Senior

XII. Where do you live? (circle one)

1. On campus 4. Own home
2. In an apartment 5. Other (please specify)
3. With parents _____

APPENDIX B

NUMBER OF PARTICIPANTS RESPONDING TO EACH CATEGORY OF TWELVE DEMOGRAPHIC TYPE ITEMS

Sex	Male	161		
	Female	72		
Age	19-21 years	4	31-40 years	13
	22-24 years	172	41 or over	2
	25-30 years	36	Unknown	6
Marital Status	Single	203		
	Married	26		
	Unknown	4		
Student Status	Full-time	228		
	Part-time	5		
Year in School	Sophomore	97		
	Junior	104		
	Senior	31		
	Grad. Student	1		
Place of Residence	Dorm	93		
	Apartment	70		
	With parents	55		
	Own home	9		
	Other	6		

Current Grade Point Average

Less than 2.0	7
2.00 - 2.50	73
2.51 - 3.00	82
3.01 - 3.50	49
3.51 - 4.00	19
Unknown	3

Estimated Final Grade Point Average at Graduation

2.00 - 2.50	27
2.51 - 3.00	90
3.01 - 3.50	95
3.51 - 4.00	19
Unknown	2

Estimated Final Grade Point Average For This Class

A	110
B	99
C	22
D	1
Unknown	1

Number of Friends Regularly Associated With

None	2
One-Two	7
Three-Five	47
Six-Eight	41
Nine-Twelve	60
Thirteen or more	75
No Response	1

Number of Clubs Currently A Member Of

None	75
One	46
Two	50
Three	34
Four	12
Five or More	13
No Response	3

Number of Elected Offices Held

None	117
One	33
Two	36
Three	22
Four	9
Five or More	16

APPENDIX C

FREQUENCY AND PERCENTAGE OF PARTICIPANTS RESPONDING TO EACH CATEGORY OF EVERY ITEM OF THE EDWARDS PERSONAL PREFERENCE SCHEDULE

	Inconsistent Responses "0"		Inconsistent Responses "1"		Missing Values	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Item 1	170	73.0	63	27.0	0	0.0
Item 2	60	25.8	172.	73.8	1	0.4
Item 3	140	60.1	92	39.5	1	0.4
Item 4	94	40.3	139	59.7	0	0.0
Item 5	201	86.3	30	12.9	2	0.9
Item 6	177	76.0	56	24.0	0	0.0
Item 7	128	54.9	103	44.2	2	0.9
Item 8	151	64.8	80	34.3	2	0.9
Item 9	157	67.4	75	32.2	1	0.4
Item 10	56	24.0	175	75.1	2	0.9
Item 11	95	40.0	138	59.2	0	0.0
Item 12	161	69.1	70	30.0	2	0.9
Item 13	152	65.2	81	34.8	0	0.0
Item 14	153	65.7	77	33.0	3	1.3
Item 15	61	26.2	167	71.7	5	2.1
Item 16	219	94.0	13	5.6	1	0.4
Item 17	142	60.9	90	38.6	1	0.4
Item 18	130	55.8	102	43.8	1	0.4
Item 19	133	57.1	100	42.9	0	0.0
Item 20	170	73.0	62	26.6	1	0.4
Item 21	109	46.8	122	52.4	2	0.9
Item 22	184	79.0	47	20.2	2	0.9
Item 23	185	79.4	47	20.2	1	0.4
Item 24	207	88.8	26	12.2	0	0.0
Item 25	169	72.5	64	27.5	0	0.0
Item 26	112	48.1	118	50.6	3	1.3
Item 27	79	33.9	153	65.7	1	0.4
Item 28	126	54.1	102	43.8	5	2.1
Item 29	150	64.4	74	31.8	9	3.9
Item 30	204	87.6	26	11.2	3	1.3
Item 31	66	28.3	158	67.8	9	3.9
Item 32	125	53.6	97	41.6	11	4.7
Item 33	138	59.2	87	37.3	8	3.4
Item 34	122	52.4	97	41.6	14	6.0
Item 35	88	37.8	111	47.6	34	14.6
Item 36	189	81.1	43	18.5	1	0.4
Item 37	177	76.0	50	21.5	6	2.6
Item 38	102	43.8	127	54.5	4	1.7
Item 39	85	36.8	144	61.8	4	1.7
Item 40	164	70.4	67	28.8	2	0.9

	Inconsistent Responses "0"		Inconsistent Responses "1"		Missing Values	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Item 41	106	45.5	124	53.2	3	1.3
Item 42	209	89.7	24	11.3	0	0.0
Item 43	185	79.4	46	19.7	2	0.9
Item 44	116	49.8	115	49.4	2	0.9
Item 45	102	43.8	129	55.4	2	0.9
Item 46	96	41.2	135	57.9	2	0.9
Item 47	56	24.0	175	75.1	2	0.9
Item 48	172	73.8	60	25.8	1	0.4
Item 49	157	67.4	74	31.8	2	0.9
Item 50	124	53.2	106	45.5	3	1.3
Item 51	116	49.8	113	48.5	4	1.7
Item 52	114	48.9	115	49.4	4	1.7
Item 53	172	73.8	57	24.5	4	1.7
Item 54	98	42.1	129	55.4	6	2.6
Item 55	142	60.9	89	38.2	2	0.9
Item 56	129	55.4	91	39.1	13	5.6
Item 57	156	67.0	74	31.8	3	1.3
Item 58	161	69.1	47	20.2	25	10.7
Item 59	90	38.6	119	51.1	24	10.3
Item 60	57	24.5	151	64.8	25	10.7
Item 61	174	74.7	57	24.5	2	0.9
Item 62	37	15.9	195	83.7	1	0.4
Item 63	200	85.8	29	12.4	4	1.7
Item 64	104	44.6	122	52.4	7	3.0
Item 65	147	63.1	81	34.8	5	2.1
Item 66	149	63.9	58	24.9	26	11.2
Item 67	200	85.8	30	12.9	3	1.3
Item 68	69	29.6	138	59.2	26	11.2
Item 69	141	60.5	89	38.2	3	1.3
Item 70	158	67.8	72	30.9	3	1.3
Item 71	160	68.7	69	29.6	4	1.7
Item 72	164	70.4	67	28.8	2	0.9
Item 73	122	52.4	94	40.3	17	7.3
Item 74	165	70.8	65	27.9	3	1.3
Item 75	131	56.2	89	38.2	13	5.6
Item 76	159	68.2	71	30.5	3	1.3
Item 77	166	71.2	65	27.9	2	0.9
Item 78	80	34.3	145	62.2	8	3.4
Item 79	184	79.0	44	18.9	5	2.1
Item 80	201	86.3	24	10.3	8	3.4

	Inconsistent Responses "0"		Inconsistent Responses "1"		Missing Values	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Item 81	155	66.5	72	30.9	6	2.6
Item 82	102	43.8	105	45.1	26	11.2
Item 83	147	63.1	77	33.0	9	3.9
Item 84	115	49.4	93	39.9	25	10.7
Item 85	147	63.1	80	34.3	6	2.6
Item 86	67	28.8	139	59.7	27	11.6
Item 87	175	75.1	54	23.2	4	1.7
Item 88	203	87.1	23	9.9	7	3.0
Item 89	159	68.2	71	30.5	3	1.3
Item 90	126	54.1	97	41.6	10	4.3
Item 91	127	54.5	102	43.8	4	1.7
Item 92	156	67.0	71	30.5	6	2.6
Item 93	190	81.5	39	16.7	4	1.7
Item 94	65	27.9	161	69.1	7	3.0
Item 95	179	76.8	48	20.6	6	2.6
Item 96	142	60.9	69	29.6	22	9.4
Item 97	209	89.7	17	7.3	7	3.0
Item 98	141	60.5	77	33.0	15	6.4
Item 99	57	24.5	166	71.2	10	4.3
Item 100	126	54.1	92	39.5	15	6.4

APPENDIX D

ITEM MEANS, STANDARD DEVIATION, AND THE PERCENTAGE OF PARTICIPANTS RESPONDING TO EACH CATEGORY OF THIRTY-SIX REACTION QUESTIONNAIRE ITEMS

	Number of subjects	Strongly Disagree 1	Disagree 2	Tend to Disagree 3	Uncertain 4	Tend to Agree 5	Agree 6	Strongly Agree 7	$\bar{X}$	σ
Item 1	232	3.0	6.0	6.0	6.5	24.6	43.5	10.3	5.16	1.47
Item 2	230	3.9	9.1	8.7	34.3	21.7	16.5	5.7	4.33	1.45
Item 3	232	0.4	1.3	1.7	5.2	15.9	49.6	25.9	5.87	1.04
Item 4	231	0.9	1.3	2.2	10.0	19.0	40.7	26.0	5.71	1.18
Item 5	231	1.3	3.0	8.2	8.7	19.9	39.0	19.9	5.39	1.39
Item 6	230	0.9	5.2	11.3	10.0	27.8	39.1	5.7	4.99	1.34
Item 7	233	7.0	5.2	4.3	7.8	18.3	44.3	13.0	5.10	1.67
Item 8	232	2.6	4.3	10.8	9.5	21.1	38.4	13.4	5.11	1.50
Item 9	227	0.9	2.2	3.5	5.3	12.8	53.3	22.0	5.75	1.18
Item 10	232	6.5	16.8	24.1	7.3	14.2	22.4	8.6	4.08	1.82
Item 11	229	1.7	5.2	16.2	16.2	23.1	29.3	8.3	4.75	1.46
Item 12	231	4.3	11.7	5.6	13.4	14.7	23.4	26.8	5.00	1.85
Item 13	232	1.3	1.3	2.2	1.3	8.6	40.5	44.8	6.16	1.14
Item 14	232	0.9	1.7	6.0	10.3	16.4	37.5	27.2	5.61	1.32
Item 15	231	0.0	1.7	3.5	15.6	29.9	38.5	10.8	5.33	1.08
Item 16	231	1.3	0.0	1.7	23.2	11.7	36.8	25.1	5.55	1.26
Item 17	231	0.0	1.7	4.8	5.2	17.3	51.9	19.0	5.70	1.09
Item 18	231	1.3	1.3	6.5	12.1	20.8	44.2	13.9	5.38	1.25
Item 19	233	6.0	6.0	20.6	11.2	27.0	19.3	3.9	4.15	1.62
Item 20	233	3.0	15.0	25.8	18.9	22.7	11.2	3.4	3.91	1.45

	Number of subjects	Strongly Disagree 1	Disagree 2	Tend to Disagree 3	Uncertain 4	Tend to Agree 5	Agree 6	Strongly Agree 7	$\bar{X}$	σ
Item 21	229	3.9	7.9	11.4	55.0	12.2	7.4	2.2	3.95	1.18
Item 22	233	0.9	3.9	7.3	13.3	29.2	36.5	9.0	5.12	1.28
Item 23	231	9.5	13.9	23.4	10.0	30.7	7.4	5.2	3.81	1.64
Item 24	232	2.6	6.0	12.9	15.5	32.3	25.4	5.2	4.66	1.41
Item 25	233	5.6	6.9	18.5	19.7	27.0	17.2	5.2	4.28	1.52
Item 26	233	3.0	15.9	24.9	9.4	26.2	18.0	2.6	4.04	1.54
Item 27	232	15.9	25.0	21.6	25.4	8.2	3.0	0.9	2.97	1.37
Item 28	232	12.9	17.2	28.4	17.7	11.6	9.9	2.2	3.36	1.57
Item 29	233	4.7	16.3	24.0	12.0	21.0	18.5	3.4	3.97	1.61
Item 30	232	0.0	2.2	2.6	12.9	25.0	42.7	14.7	5.47	1.11
Item 31	232	1.7	2.2	4.7	60.3	14.7	13.4	3.0	4.36	1.06
Item 32	233	0.9	2.6	8.6	7.3	36.1	39.9	4.7	5.14	1.17
Item 33	232	0.9	3.0	7.8	44.8	23.3	17.2	3.0	4.50	1.11
Item 34	231	8.2	6.1	11.7	15.6	29.9	23.8	4.8	4.43	1.61
Item 35	232	0.0	0.9	5.6	21.6	25.4	34.9	11.6	5.23	1.13
Item 36	232	8.2	10.3	13.8	23.7	27.2	13.8	3.0	4.05	1.55

APPENDIX E

PERCENTAGE OF PARTICIPANTS RESPONDING TO EACH OF FOUR CATEGORIES FOR EACH OF FORTY STATE-TRAIT ANXIETY INVENTORY ITEMS

	Not At All	Some-what	Moderately So	Very Much So	Missing Responses
Item 1	1.3	15.9	52.2	30.6	1
Item 2	0.4	14.7	49.1	35.8	1
Item 3	60.3	29.7	7.8	2.2	1
Item 4	70.6	18.6	8.7	2.2	2
Item 5	1.7	19.4	44.4	34.5	1
Item 6	76.6	15.6	6.9	0.9	2
Item 7	39.0	38.5	13.4	9.1	2
Item 8	16.9	40.3	30.7	12.1	2
Item 9	36.8	32.9	22.9	7.4	2
Item 10	5.6	25.0	48.3	20.7	2
Item 11	2.2	19.0	52.2	26.7	1
Item 12	72.0	20.3	5.6	2.2	1
Item 13	75.8	18.6	3.9	1.7	2
Item 14	64.1	24.2	8.2	3.0	3
Item 15	1.3	31.2	46.3	21.2	2
Item 16	7.9	24.5	50.7	17.0	4
Item 17	51.9	31.6	10.8	5.6	2
Item 18	78.8	13.4	6.5	1.3	2
Item 19	17.2	41.4	31.9	9.5	1
Item 20	6.1	30.0	48.7	15.2	3
Item 21	0.4	12.5	66.8	20.3	1
Item 22	28.9	59.1	9.9	2.2	1
Item 23	50.9	43.1	5.6	0.4	1
Item 24	28.5	52.6	13.6	5.3	5
Item 25	49.6	39.6	9.6	1.3	3
Item 26	4.8	35.1	43.3	16.9	2
Item 27	0.9	28.6	49.8	20.8	2
Item 28	30.3	55.0	13.9	0.9	2
Item 29	22.6	47.4	22.2	7.8	3
Item 30	0.4	15.7	50.4	33.5	3
Item 31	6.5	57.1	26.8	9.5	2
Item 32	35.5	48.5	11.7	4.3	2
Item 33	0.9	21.7	50.0	27.4	3
Item 34	20.3	46.8	20.8	12.1	2
Item 35	34.6	59.3	6.1	0.0	2
Item 36	2.2	20.4	54.3	23.0	3
Item 37	24.2	53.2	16.5	6.1	2
Item 38	31.3	49.6	15.2	3.9	3
Item 39	1.7	19.8	51.3	27.2	1
Item 40	31.6	51.3	12.7	4.4	5

APPENDIX F

LOADINGS OF THIRTY-SIX REACTION QUESTIONNAIRE ITEMS ON EACH OF FIVE REACTION FACTORS

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Item 1	.17	-.09	-.15	-.06	.01
Item 2	.37	-.27	-.12	-.02	.09
Item 3	.18	-.11	.02	-.10	-.03
Item 4	-.02	.25	.13	.02	-.02
Item 5	-.07	.12	.09	.15	-.05
Item 6	-.04	.03	-.17	-.15	.11
Item 7	.07	-.04	-.21	.06	.12
Item 8	-.04	.23	.48	-.01	.06
Item 9	.07	-.20	-.37	-.04	.09
Item 10	-.02	.03	.77	.04	.03
Item 11	.11	-.03	-.61	-.10	.10
Item 12	-.09	.13	.08	-.07	-.14
Item 13	-.02	-.10	.01	-.01	.07
Item 14	-.12	.15	.08	.01	-.41
Item 15	.12	-.16	-.01	-.07	.88
Item 16	-.16	.45	.14	-.06	-.11
Item 17	.09	-.14	-.04	-.03	.08
Item 18	-.07	.63	-.06	-.02	-.10
Item 19	.50	-.03	-.09	-.23	.19
Item 20	.50	-.16	-.10	-.11	.01
Item 21	.61	-.16	-.09	-.02	.13
Item 22	-.16	.53	.15	.10	.00
Item 23	-.36	.24	.12	.38	-.01
Item 24	-.28	.39	-.03	.22	-.10
Item 25	-.33	.26	.28	.44	-.08
Item 26	.13	.08	.06	-.10	-.06
Item 27	-.14	.12	-.03	.16	-.08
Item 28	-.06	.05	.11	.70	-.06
Item 29	-.01	.14	.03	.11	.01
Item 30	-.11	.39	.08	.06	-.13
Item 31	.70	-.06	.03	-.02	-.00
Item 32	.32	-.13	-.07	-.05	.06
Item 33	.73	.02	-.08	-.00	.08
Item 34	.33	-.19	-.13	-.19	.09
Item 35	.13	.42	.25	.16	-.08
Item 36	.37	-.20	-.07	-.13	.12

VITA

Kenneth A. Horner was born in Rockford, Illinois on December 11, 1951. He attended public schools in that city and graduated from Harlem High School, Loves Park, Illinois in June of 1969. He attended Rock Valley Junior College, Rockford, Illinois prior to entering the United States Air Force in March of 1971. While in the Air Force he attended the Electronic Intelligence Technicians School at Keesler Air Force Base (AFB), Mississippi and served tours of duty in Alaska and at Strategic Air Command Headquarters, Offut AFB, Nebraska. After receiving an honorable discharge from the Air Force in December of 1974, he attended Rockford College, Rockford, Illinois where he received a Bachelor of Science degree in Psychology in May, 1977. The following September he entered the Graduate School of The University of Tennessee. He received a Master of Science degree in Organizational Psychology in August of 1979.