The Effects of Test Anxiety and Test Conditions upon Consistency in Cognitive Functioning: A Study of Intra-Test Scatter and Internal Consistency on the WISC Comprehension Subtest

Lawrence H. Zaiden
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Education Commons

Recommended Citation
https://ecommons.luc.edu/luc_diss/1239

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 1972 Lawrence H. Zaiden
The Effects of Test Anxiety and Test Conditions
Upon Consistency in Cognitive Functioning:
A Study of Intra-test Scatter and
Internal Consistency on the
WISC Comprehension Subtest

by

Lawrence H. Zaiden

A Dissertation Submitted to the Faculty of the School of Education
of Loyola University in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

June
1972
ABSTRACT

The present study was designed to investigate the effects of test anxiety and test conditions upon consistency in intelligence test performance. On the basis of previous research, it was assumed that anxiety in test situations, whether predispositional or reactive in nature, interferes with efficiency in cognitive performance. On this basis, the following hypotheses were proposed:

1. High anxious Ss, regardless of whether they are tested under either stressful or nonstressful conditions, will show a significantly higher degree of inconsistency in their intellectual test performance than low anxious Ss.

2. Those Ss tested under stressful conditions, regardless of whether they have been classified as high or low anxious, will show a significantly higher degree of inconsistency in their intellectual test performance than those Ss tested under nonstressful conditions.

3. There will be a significant interaction effect between stress and prior test anxiety level. Accordingly, the highest degree of inconsistency in intellectual performance will be seen in the high anxious subgroup tested under stressful conditions. Conversely, the lowest degree of inconsistency in intellectual performance will be seen in the low anxious subgroup tested under nonstressful conditions.

The sample consisted of 96 fifth-graders with average I.Q.'s attending public school in a predominantly white, middle-class, suburban region. A standardized questionnaire (TASC) was used to select LA and HA Ss. Ss were paired on the basis of sex, age, group I.Q. score, and test anxiety level.
One member of each pair was then randomly assigned to a nonstressful treatment condition; the other member to the stressful treatment condition. The intelligence test used was the Comprehension subscale of the WISC. Half of the 48 LA Ss were individually tested under nonstressful conditions; the other half, under stressful treatment conditions. A similar procedure was followed with the remaining 48 HA Ss.

Under the null hypothesis, $H_0: \rho = 0$, the internal-consistency reliability measures revealed that the high anxious Ss performed with a greater degree of inconsistency in their item-to-item responding than the low anxious Ss, when the effects of test conditions were partialed out. Second, stressed Ss performed with a greater degree of inconsistency than nonstressed Ss, when the effects of predispositional anxiety level were partialed out. Third, induced anxiety, as evidenced by the marked discrepancy in consistency of responding between the high anxious, stressed Ss and the low anxious, non-stressed Ss. As anticipated, these differences were obtained irrespective of any differences in level of performance. Hence, these results support the experimental hypotheses. However, under the null hypothesis, $H_0: \tau_1 - \tau_2 = 0$, the difference in the internal-consistency reliability measures were not found to be discrepant enough to be considered significant at conventional levels of significance. In addition, two clinically derived measures of intra-test scatter were found to be ineffective measures of intra-individual response consistency.

The results suggest the need for additional research to further determine the effects of personality variables and situational variables upon intra-individual response consistency in intellectual test performance.
ACKNOWLEDGEMENTS

The writer wishes to express his appreciation to Dr. Samuel Mayo, Advisor, to Dr. Anne Juhasz, and to Dr. Henry Malecki for their thoughtful suggestions and encouragement which aided in the preparation of this dissertation.

Gratitude is also extended to Mr. William Young, Director, and to the board members and staff of the School Association for Special Education in Du Page for their consideration and support in furthering this effort.

The administration and staff of the Glendale Heights, Bensenville, and Lisle elementary school districts were also extremely helpful and cooperative in providing subjects and a setting for the study.

Finally, the writer wishes especially to express his gratitude to his wife, Eleanore, who contributed so vitally to this dissertation through her assistance, understanding, and encouragement.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF THE LITERATURE</td>
<td>13</td>
</tr>
<tr>
<td>III. METHOD</td>
<td>34</td>
</tr>
<tr>
<td>IV. RESULTS</td>
<td>46</td>
</tr>
<tr>
<td>V. DISCUSSION</td>
<td>67</td>
</tr>
<tr>
<td>References</td>
<td>86</td>
</tr>
<tr>
<td>Appendix A Forms</td>
<td>93</td>
</tr>
</tbody>
</table>
TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ANOVA on Matching of Ss on the Basis of C.A., according to Sex, Assignment to Treatment Condition, and Anxiety Level</td>
<td>37</td>
</tr>
<tr>
<td>2. ANOVA on Matching of Ss on the Basis of I.Q. according to Sex, Assignment to Treatment Condition, and Anxiety Level</td>
<td>38</td>
</tr>
<tr>
<td>3. ANOVA: Differences in Total Score Associated with Sex, Test Conditions, and Anxiety Level</td>
<td>49</td>
</tr>
<tr>
<td>4. Sex Differences in Internal Consistency as Measured by the Correlation Between the Odd-and Even-Numbered Items ($r_{ppm}$), and the Spearman-Brown ($r_{SB}$) and Kuder-Richardson ($r_{KR20}$) Formulas</td>
<td>54</td>
</tr>
<tr>
<td>5. Differences in Internal Consistency Associated with Anxiety Level as Measured by the Correlation Between the Odd-and Even-Numbered Items ($r_{ppm}$), and the Spearman-Brown ($r_{SB}$) and Kuder-Richardson ($r_{KR20}$) Formulas</td>
<td>55</td>
</tr>
<tr>
<td>6. Differences in Internal Consistency Associated with Test Conditions as Measured by the Correlation Between the Odd-and Even-Numbered Items ($r_{ppm}$) and the Spearman-Brown ($r_{SB}$) and Kuder-Richardson ($r_{KR20}$) Formulas</td>
<td>57</td>
</tr>
<tr>
<td>7. Differences in Internal Consistency Associated with the Interaction Between Anxiety and Test Conditions as Measured by the Correlation Between the Odd-and Even-Numbered Items ($r_{ppm}$) and the Spearman-Brown ($r_{SB}$) and Kuder-Richardson ($r_{KR20}$) Formulas</td>
<td>59</td>
</tr>
<tr>
<td>8. ANOVA: Differences in Intra-test Scatter Associated with Sex, Test Conditions, and Anxiety Level with Runs Used as the Criterion Measure (Responses Scored Pass or Fail)</td>
<td>63</td>
</tr>
<tr>
<td>9. ANOVA: Differences in Intra-test Scatter Associated with Sex, Test Conditions, and Anxiety Level with Runs Used as the Criterion Measure (Response Scored 2, 1, or 0)</td>
<td>65</td>
</tr>
</tbody>
</table>
Anxiety has become an important topic for research within the behavioral sciences, particularly within the fields of education and psychology. From a theoretical standpoint, its relationship to cognitive functioning has been a topic of special importance. For example, few personality theorists and developmental psychologists have failed to stress the influence of anxiety upon the affective and cognitive development of the individual. Learning theorists, too, have shown an increasing interest in more dynamic theories of learning which take account of anxiety. Anxiety, variously defined, has been given a central role in many contemporary learning theories, such as those of Spence and Spence, Mowrer, and Dollard and Miller (Ruebush, 1963).

To the pragmatically oriented educator, the relationship between anxiety and intelligence is an especially pertinent topic. In this connection, the persistent problem of underachievement is one of great importance and one in which anxiety is viewed as being intimately involved (Dudek & Lester, 1968; Roth, Mauksch & Peiser, 1967; Roth & Meyersburg, 1965; S. B. Sarason, Davidson, Lighthall, Waite, & Ruebush, 1960). This problem is one of considerable concern in our schools today, since it is commonly recognized that there is a large number of youngsters who fail to realize their potentials within the school environment.

Because anxiety has been a topic of research in so many diverse areas, efforts to integrate empirical findings with theory have been fraught with difficulty. In particular, the matter of definition has proved quite trouble-
some. This difficulty is commented upon by English and English (1958) as follows:

When a term is frequently employed in behavioristic learning theory, in psychoanalysis, and in nearly every field of psychology between them, the variety and shadings of meaning become very troublesome. Anxiety must be read with great vigilance for an author's meaning or, more often than not, his several meanings (p. 35).

In spite of the many varied meanings attached to anxiety, some commonalities in definition may be found. In general, most authors regard anxiety as a vaguely unpleasant feeling of dread or anticipation of possible harm resembling a secondary drive, which tends to evoke an avoidance response. This condition may be accompanied by observable physiological changes, somatic concerns, and/or measurable effects upon task performance (English & English, 1958; Klausmeier & Goodwin, 1966; McReynolds, 1968; Phillips, Martin, & Meyers, 1970; Ruebush, 1963; S. B. Sarason, et al., 1960; Spielberger, 1966b).

Despite the diversity in commonly used general definitions of anxiety, McReynolds (1968) concludes, "It can be said that the construct of anxiety, though conceptually imprecise, has nevertheless proved generally useful, and seems basically well founded (p. 247)."

In an attempt to render the concept more directly applicable to empirical investigation, McReynolds notes that many authors, including those cited above, have implicitly treated anxiety as a multidimensional construct. For example, along one dimension, anxiety may be dichotomized as "characteristic (trait) anxiety" versus "current (state) anxiety." Along another dimension, it is appropriate to speak of "existent anxiety" versus "prone-ness to anxiety", according to McReynolds (p.263). This multidimensional conceptualization of anxiety adds some clarity and will be delineated in somewhat greater detail in the next chapter. However, at this point, it is meaningful to note especially the distinction between "general" and "test" anxiety, as advanced by one experienced investigator and his colleagues...
The concept of test anxiety is one which has proved especially pertinent to the field of educational and psychological measurement, the field in which the focus of this investigation will lie. In brief, test anxiety, may be regarded as a predisposition to feel anxious in test and test-like situations. Of specific interest is the notion that under certain conditions test anxiety will interfere with intellectual performance. S. B. Sarason and his associates presented empirical support for this notion. However, Sarason and other investigators point up the need for further research to determine more precisely how, under what circumstances, and in what ways such interfering effects will be manifested. Specific attention needs to be directed not only to subject variables (e.g., anxiety level) but also to situational variables (i.e., the conditions under which the task is administered).

Comparatively little attention has been given to the problem of determining in what specific ways the interfering effects of test anxiety may be manifested in test performance. In the main, previous investigations have focused primarily upon level of performance as the dependent variable. In other words, most investigators have sought only to determine whether or not test anxiety, under varying conditions, may be associated with a general or overall performance decrement, such as a reduced overall intelligence test score. However, comparatively few empirical investigations have dealt with other consequents of test anxiety which may be seen in varying patterns of test performance. For example, no studies could be located which have dealt exclusively with the matter of consistency in performance as it relates to the topic of test anxiety. Yet the matter is an important one; for, as will be seen later, to the extent that an examinee can consistently utilize his
abilities in item-to-item responding, to that extent do our general notions about the reliability and validity of the test apply.

Viewed from a clinical point of view, consistency in cognitive task performance is looked upon as being an important indication of intellectual efficiency, which is associated with freedom from handicapping affective difficulties, such as disruptive anxiety. Conversely, the degree of inconsistency (or "scatter" as it is often referred to in conjunction with test performance) is often associated with the degree of affective disturbance which is present. This view was most notably put forth by Rapaport, Gill, and Schafer (1945) and by Wechsler (1958) and is widely supported by clinicians.

In relation to point scales (i.e., tests composed of items graded in difficulty from easy to hard as are the Wechsler scales), these authors have suggested that disruptions in performance associated with affective disturbance may be evident from the amount of intra-test scatter in an examinee's performance. On such cognitive tests, intra-test scatter is usually shown when the testee fails easy and passes hard items within the same scale.

The implications of this variable response pattern may perhaps be better understood if dealt with in concrete terms. As an example, two examinees may obtain identical total scores on a given point scale. The performance of one examinee may reflect a very high degree of consistency and efficiency in item-to-item responding. If the items are graded in difficulty, ideally, this individual would be expected to obtain credit for all items until he reaches his threshold of difficulty, or point at which the items become so difficult for him that no more items are passed. In contrast, the other examinee might conceivably perform in a far less consistent and less efficient fashion than his counterpart, even though he attains the same total score on the test. This erraticism may be evident from the degree of intra-test scatter in his item-to-item responding. If this examinee was known to be
test anxious, one might hypothesize that he might really be capable of attaining a higher score but for the disruptive effects of anxiety.

While these notions have their origins in the field of abnormal and clinical psychology, it is applicable to the extent that a clear-cut distinction between "normal" and "abnormal" behavior may be made. The history of psychology is replete with examples in which clinical studies of abnormal behavior has furthered an understanding of normal behavior.

Wechsler (1958), in keeping with the traditional view that abnormal behavioral patterns may also be observed under certain conditions in "normal" individuals, makes the following comments regarding the interfering effects of anxiety:

Anxiety is a state that occurs in many mental disorders and is not necessarily diagnostic of any one of them. As a transient manifestation it may be observed in normal persons when under severe stress or threat.

Anxiety may be either a cause or symptom. In either case it is generally disruptive or disabling. It may interfere with the individual's functioning in a variety of ways. On the physical side it may reflect itself in tremor, restlessness, physiological inhibition, dryness of mouth, palpitation, etc.; mentally, by inability to concentrate, fluctuations in attention, moodiness and erraticism. In test performance any or all of these manifestations may be present (pp. 175-176).

Hence, if these assumptions are tenable, the mental abberations associated with "inability to concentrate" and with "fluctuations in attention" would conceivably result in "erraticism" in test performance.

Although these assumptions are important in psychological theory and are widely held by psychodiagnosticians, they have not been sufficiently investigated through systematic, empirical research. In the current literature there have been few attempts to relate the assumptions to research on the performance effects of anxiety. Attention to consistency in intellectual test performance has remained largely a "clinico-intuitive operation or qualitative type of operation (p. 493)", according to Rabin (1965) in his discussion on the diag-
nostic use of intelligence tests.

Statement of the Problem

The present investigation will focus upon the question of whether or not test anxiety results in intra-individual response variability or inconsistency in item-to-item responding on an intelligence test. As suggested by previous investigators, the question will be considered within the context of the actual conditions under which the test is administered. In this case, the test conditions will include a comparison of stressful versus nonstressful test administration.

Purpose of the Study and Significance of the Problem

The problem is viewed as having both theoretical and practical significance for many fields, but it is especially pertinent to the field of educational and psychological measurement. From the standpoint of test theory, inconsistency or variability in item-to-item, as it is associated with personality functioning, has been related directly to the problem of test reliability and validity (Eysenck, 1947; Fiske & Rice, 1955; Loevinger, 1967; Sechrest & Jackson, 1967; Thorndike, 1951; White & Saltz, 1967). Along these lines, Guilford (1954) discussed the relationship between consistency and level of performance. The following comments outline the significance of the problem:

There has been considerable interest in very recent years in personal variability in measured ability. It is sometimes pointed out that we should know not only the examinee's characteristic level on a scale of ability but also his degree of consistency in performing near that level. It is possible that individuals differ systematically from one another in their consistency as well as their level of performance. If this is so, we can obtain additional, useful information about individuals. If a certain examinee is quite consistent, his level of performance will be quite predictable. If another examinee is markedly inconsistent or
or variable about his mean, he is to that extent unpredictable in this ability. Usual test practices seem to operate on the assumption that all examinees are equally predictable.

There is the larger psychological question whether such personal variability is specific to different tests or is a more general trait that would be found common to a variety of tests. One is also reminded of the clinical interest in "spread" of performance, which is usually detected over a variety of tests. There are thus intermingled problems of variability over time as well as over difficulty levels and over different abilities. The psychological meanings of none of these phenomena are as yet very clear. They represent a challenge of unsolved measurement problems. (pp. 347-348).

Most tests are standardized on "normal" subjects often with insufficient attention directed to individual differences in temperament which affect test-taking behavior and performance. Based upon standardization data, certain statistical and actuarial conclusions are drawn and applied to all individuals who will be given the test. However, as Sechrest and Jackson (1967) note, there are many individuals who often do not fall within the normal or expected pattern of response. This sometimes happens for reasons which cannot be accounted for simply in terms of the deficiency or lack of precision of the measuring instrument. For example, when parallel forms of a test are given, often the scores of some individuals on one form tend for some reason to be unrelated to their scores on the parallel form of the test. Such individuals are referred to as correlational "outliers", for their scores tend to fall outside the regression pattern obtained for all scores. Sechrest and Jackson emphasize the need to explain why such individuals tend to deviate. Perhaps the test anxiety hypothesis may provide one explanation; since, in view of considerations stated above, it may be hypothesized that perhaps those individuals who tend to be most anxious about tests also tend to deviate in their test performance to the greatest degree. In any event this problem would serve to illustrate the need for a greater reconciliation between measurement technology and psychological theory as called for by many authors (Anastasi,
1967; Bock & Wood, 1971; Cronbach, 1967; Mayo, 1965). This, then, would be one objective of the present research.

Aside from these theoretical considerations, research on the matter of test anxiety and its effects upon consistency in test performance has direct practical significance. As was noted above (Rabin, 1965), diagnostic interpretations based upon notions regarding emotional stability and consistency in cognitive performance have remained largely an intuitive affair. Thus, for the diagnostician there is a need to place his test interpretations upon a more solid footing. With respect to the problem of underachievement referred to above, perhaps more underachievers could be more adequately identified and provided with help if there were a greater variety of empirically based diagnostic techniques available. This matter is succinctly dealt with by Sarason, et al. (1960), as reflected in the following comments:

When one considers the mass manner in which tests are frequently given, precluding the possibility of observing test behavior, it would seem to be only fair (a not unimportant consideration) also to obtain other data which might put the test performance in another light. In these days when our society seems to have become aware of the problems of creativity, wasted talent and early diagnosis of personality disturbance, the psychologist must seriously consider methodologies which, on a priori grounds, give some hope of sharpening our diagnostic and predictive statements about individuals. The practice in the past has been to increase the level of validity of prediction by developing more and new tests, a practice which in itself is certainly appropriate. However, in the light of what we have discussed above, such a practice would seem to make the study of test-taking reactions even more pressing (p. 9).

Along these lines high anxious individuals, who as hypothesized would tend to be variable in their test performance, would be at a disadvantage in school settings where the tendency is to over-emphasize total score differences for classification or selection purposes. Certainly inappropriate use of psychological tests has been a matter of much concern by authors of texts on tests and measurement (e.g., Cronbach, 1970, pp. 22-35; Nunnally, 1964, pp. 10-13). These concerns have been intensified by recent attacks upon psychological
testing, as discussed by Anastasi (1967) and more recently by Kirkland (1971). Hence, there is a clear-cut need for more research dealing with the matter of individual differences as they may be manifested in divergent patterns of test performance.

Further, research of this nature has significance and applicability in addition to that directly related to the field of educational and psychological measurement. Some of these implications have been stated above as they pertain to anxiety-related research. For example, the question of cause and effect is basic to theories dealing with the relationship between anxiety and intelligence. Thus, the more we know of anxiety through empirical research, the more adequately we can relate such conclusions to theories of personality development, learning, psychopathology and to other related areas.

Limitations

The present investigation will deal with the relationships among test anxiety, test conditions, and consistency of intelligence test performance in elementary school children. But before stating the experimental hypotheses, it is necessary to specify some of the major limitations restricting the scope of this investigation.

The term, "anxiety", will be limited in the sense that it will be used primarily with reference to what has been specifically described as "test anxiety." Operationally, for the purpose of this research, test anxiety as a psychological construct is limited to the degree of test-related anxiety which the subject will admit to in response to a standardized questionnaire, the Test Anxiety Scale for Children (TASC). Hence, where the term, "anxiety", is used in this investigation, it will refer specifically to "test anxiety" as measured by the TASC.
By the same token, consideration of "intelligence" as a psychological construct will be operationally limited to the actual measures employed in this experiment. Attention will focus upon the Comprehension subtest of the WISC, mainly in view of its special sensitivity to vicissitudes in personality functioning (Glasser & Zimmerman, 1967; Ogdon, 1967). As a whole, the WISC has been favorably reviewed by recent reviewers (Burstein, 1965; Fraser, 1959; Littell, 1960; Patterson, 1959; Rabin, 1959) and has been widely studied and accepted. Specifically, interest will center upon consistency in performance as it relates to item-to-item responding. Ss will be compared on several consistency measures (dependent variable) rather than solely upon level of performance. The rationale for this procedure has been outlined above and will be elaborated upon in the next section.

A third major limitation of this investigation is, of course, the population to which the findings may be logically and statistically generalized. Interest is focused upon children in the middle or intermediate grades. It has been widely asserted that this is an important and difficult transitional period in a youngster's school experience and a point at which anxieties about tests and test-like situations are very much in evidence (Kagan & Kagan, 1970; Klausmeier & Goodwin, 1966; S. B. Sarason, et al., 1960). The characteristics of the sample selected for this study are described in another section of this paper.

Another limitation of this investigation relates to the extent to which the specific experimental procedures employed may be generalized. In the present study, two treatment comparisons will be made. Ss tested under relatively stressful conditions will be compared with Ss tested under non-stressful conditions. The actual procedures employed and the rationale will be described in a later chapter.
Hypotheses

On the basis of previous research and in view of theoretical considerations outlined above, it is predicted that test anxiety will have a detrimental effect upon consistency in intellectual test performance. However, the effects of test anxiety will be investigated in relation to the differential treatment variable, test conditions. Specifically, the following hypotheses are presented:

1. High anxious Ss, regardless of whether they are tested under either stressful or nonstressful conditions, will show a significantly higher degree of inconsistency in their intellectual test performance than low anxious Ss.

2. Those Ss tested under stressful conditions, regardless of whether they have been classified as high or low anxious, will show a significantly higher degree of inconsistency in their intellectual test performance than Ss tested under nonstressful conditions.

3. There will be a significant interaction effect between stress and prior test anxiety level. Accordingly, the highest degree of inconsistency in intellectual performance will be seen in the high anxious subgroup tested under stressful conditions. Conversely, the lowest degree of inconsistency in intellectual performance will be seen in the low anxious subgroup tested under nonstressful conditions.

Summary and Overview

In this chapter, test anxiety, as an explanatory concept, was introduced and differentiated from more general notions regarding anxiety. The relation-
ship among test anxiety, test conditions, and intelligence test performance was then discussed by way of brief introduction. Interest was directed specifically to consistency in intelligence test performance, which was described as a dependent variable that has not been widely studied. The theoretical ramifications and practical applications of this research were outlined briefly. The problem was then stated along with the purpose and limitations of this investigation. Finally, the research hypotheses were presented.

Chapter II will present a review of the related literature. A more extensive discussion of the problem under investigation will be presented along with its ramifications. Chapter III will describe the methodology and procedures used in selecting and matching the Ss. The results will be presented in Chapter IV, followed by a discussion of the implications of this research to be presented in Chapter V.
CHAPTER II

REVIEW OF THE LITERATURE

The initial task of this review will be to more fully delineate and to integrate the theoretical assumptions underlying the present investigation. Following a theoretical discussion of the construct of anxiety, interest will focus specifically upon the concept of test anxiety. Some of the major empirical findings on the effects of test anxiety upon intellectual test performance will then be presented. These findings will be compared and contrasted in order to determine in what areas additional research is needed. Finally, the chapter will conclude with a discussion of some of the ramifications of research of this nature.

Anxiety as a Psychological Construct

That anxiety has proved to be a fruitful topic for educational and psychological research is evidenced in part by the massive literature which it has generated over the years. According to Spielberger (1966b), anxiety research received considerable impetus since the early 1950's. This period was marked by attempts to apply psychoanalytically derived notions regarding anxiety to other fields, such as learning theory. Spielberger indicates that for the fifteen-year period since 1950, alone, there were over 3500 publications. Obviously, such a mass of divergent literature could not be adequately reviewed within the scope of the present paper.

Recent, comprehensive reviews of anxiety-related research and theory were cited in Chapter I. These include publications by McReynolds (1968),
Spielberger (1966a, 1966b), Ruebush (1963), Phillips, et al., (1970), and S. B. Sarason, (1960). The latter publication deals most extensively with anxiety in elementary school children and has been widely cited in this paper and in other writings.

As was noted in the previous chapter, anxiety has proved to be an elusive concept because it has been so broadly studied and so widely applied. In accordance with English and English (1958) and in an attempt to integrate various theoretical points of view (e.g., Klausmeier & Goodwin, 1966; McReynolds, 1968; Phillips, et al., 1970; Ruebush, 1963; S. B. Sarason, et al., 1960; Spielberger, 1966b), anxiety was described earlier as a dysphoric, affective condition marked by an often diffuse feeling of dread or anticipation of possible harm. Its diffuse, nonspecific nature distinguishes anxiety from fear. In addition, anxiety is sometimes viewed as a warning device or as a signaling mechanism. In this sense it resembles a secondary drive, in that it tends to evoke an avoidance response or a group of avoidance behaviors. In humans the condition may be accompanied by reported somatic concerns (e.g., mutilation fantasies, imagined injury) observable physiological changes (e.g., heightened GSR, increased pulse rate) and/or measurable effects upon task performance.

This highly general definition or description of anxiety, however, would prove to be of little benefit to empirical research because of its nonspecific nature. The confusion contained in the concept of anxiety and in varied attempts to apply the concept to empirical research were discussed by Spielberger (1966b) as follows:

Ambiguity in the conceptual status of anxiety arises from the more or less indiscriminate use of the term to refer to two very different types of concepts. Anxiety is perhaps most commonly used in an empirical sense to denote a complex reaction or response -- a transitory state or condition of the organism that varies in intensity and fluctuates over time. But the term anxiety is also used to refer to a personality trait -- to individual differences in the extent to which
different people are characterized by anxiety states and by prominent defenses against such states (p. 12).

The Multidimensional Nature of Anxiety

As was mentioned earlier, in order to render the concept amenable to empirical investigation, many authors have attempted to more narrowly define anxiety. Investigators have described anxiety in dichotomous terms from several standpoints. One notable dichotomy is the trait-state distinction, which is described by Spielberger (1966b) as follows:

Empirical evidence of different types of anxiety concepts has emerged from the factor analytic studies of Cattell and Scheier (1958; 1961). These investigators identified two distinct anxiety factors which they labeled trait anxiety and state anxiety on the basis of the procedures by which these factors were isolated and the variables which loaded on them ... The trait anxiety factor was interpreted as measuring stable individual differences in a unitary, relatively permanent personality characteristic. The state anxiety factor was based on a pattern of variables that varied over occasions of measurement, defining a transitory state or condition of the organism which fluctuated over time (p. 13).

The distinction is further clarified by Spielberger with the following analogy:

The relation between the state and trait anxiety may be conceived of as analogous in certain respects to the relation between the physical concepts of kinetic and potential energy. State anxiety, like kinetic energy, refers to an empirical process or reaction which is taking place now at a given level of intensity. Trait anxiety, like potential energy, indicates a latent disposition for a reaction of a certain type to occur if it is triggered by appropriate (sufficiently stressful) stimuli (p. 16).

This two-factor theory of anxiety is partly reflected in the views adhered to by most of the authors cited above and also by others (e.g., I. G. Sarason & Smith, 1971). However, other two-part distinctions have been substituted for the trait-state dichotomy. Phillips, et al. (1970) discuss some of the implications of these conceptualizations. Their views, which serve to integrate various points of view referred to above and below, are summarized in the following comments:
1. Anxiety has a two-part conceptual status, including what is referred to as "trait," "neurotic," or "chronic" anxiety, and what is called "state," "objective," or "situational" anxiety. Trait anxiety is dispositional in nature, is construed to be a proneness to be anxious, is primarily a function of past experience, and has an internalized (i.e., intra-psychic) locus, while state anxiety is situational in nature, is directly a function of stressful conditions, and has a contemporary locus. Moreover, these two variables interact in a manner such that anxiety proneness influences the extent of the anxiety reaction.

2. Anxiety is elicited by psychological stress and stress is reflected in threatened deprivation of an anticipated satisfaction. In addition, uncertainty associated with external (i.e., environmental) and internal (i.e., cognitive) factors is a key ingredient in whether an anxiety reaction usually occurs as a response to stress in conjunction with other affects, defensiveness, other coping reactions, etc.

3. The consequences of anxiety are usually negative, interfering and debilitating, although consequences depend on the demands and requirements of tasks and situations. Therefore, the consequences are complex and interactive in nature -- and, as a result, are sometimes facilitating, helpful, and enhancing to adaptation, performance, and adjustment (pp. 3-4).

It is evident that these varied but overlapping, theoretical descriptions of anxiety suggest that the construct may be dichotomized along several dimensions. For example, the trait-state dichotomy may be looked upon as one of several dimensions. This multidimensional model is presented by McReynolds (1968), as follows:

It is important to emphasize that the assessment of anxiety does not involve merely a single variable, but is concerned with several quite distinct dimensions. This is a fact that many test users, as well as test designers, have not taken sufficiently into account. While there have been a number of attempts - after given tests are in use - to determine just what aspect(s) of anxiety they actually measure, it is obviously preferable to construct tests for specific manageable purposes in the first place. In a measurement sense there is no such simple variable as "anxiety," but only such specific variables as "characteristic anxiety," "current proneness to test anxiety," and so on. There appear to be, in principle, at least eight potential types of anxiety scores ... (though not all would necessarily be useful). These can be conceptualized ... by thinking of a cube, the dimensions of which are the dichotomies "characteristic vs. current," "overall vs. specific," and "existent vs. proneness." For example, the IPAT Anxiety Scale is a measure of characteristic, overall, existent anxiety; the TAQ is a measure of characteristic proneness to a specific (test-taking) anxiety; and so on. Further, any of the eight scores can, in principle, be assessed in terms of either stimulus or response concomitants (pp. 262-263).

While one might argue for a model which represents anxiety in more than three...
dimensions, Fiske and Pearson (1970) feel that McReynolds' simple formulation lends perspective and helps reduce some of the existing confusion in attempting to generalize from author to author.

Test Anxiety

As was noted earlier, one conceptualization which has proved especially useful for research is the distinction between general anxiety and test anxiety. This distinction was most notably put forth by S. B. Sarason, and his associates (1960). They developed separate questionnaires for measuring both test and general anxiety in terms of the situational dimension in which the anxiety condition is manifested. Their research suggested that in contrast to the pervasive character of general anxiety, test anxiety may be regarded as characteristic proneness to feel anxious within a specific (test-taking) situation (McReynolds, 1968). But before pursuing the topic further, it is necessary to examine the formal status of the test anxiety variable as it is commonly employed in research and as it relates to the present investigation.

Ruebush (1963) notes that comparisons among research findings are more easily made when anxiety (in this case, test anxiety) is viewed as a hypothetical or theoretical construct. Accordingly he states:

Used in this way, (test) anxiety is assumed to be an entity or process that actually exists (but is not at present fully observable) and which gives rise to measurable phenomena including phenomena other than the observables that led to hypothesizing the construct. Thus, anxiety, inferred from one aspect of the child's behavior (e.g., stating his feelings on a questionnaire) is conceived of having certain other, predictable consequences (e.g., interference with performance ...) which follows from the role of the construct in its theoretical framework (pp. 462-463).

As a hypothetical construct within the present context, test anxiety may be regarded as "a predispositional process variable", since according to
Ruebush, "the anxious child is not thought to become anxious in all situations but to have a predisposition to become anxious in certain situations specified by the theory (p. 463)."

A major work (S. B. Sarason, et al., 1960) was referred to earlier as being especially relevant to the present research. It is notable in that it represents an attempt to derive empirical evidence for the hypothetical construct, test anxiety, which is treated as a predispositional process variable. The investigatory hypotheses employed by these researchers are drawn largely from psychoanalytic theory. However, by narrowing the focus of their investigation upon test anxiety, Sarason and his colleagues hoped to shed light upon some of the more general aspect of anxiety.

In the view of these investigators, anxiety experienced in the test situation or in all test-like situations stems from the child's early feelings toward the parents and especially about being evaluated by them. It was further theorized that the test anxious child tends to view the parents and all parental surrogates, i.e., all adults in positions of authority, as evaluators whose expectations he must meet in order to avoid real or imagined punishment. Primarily because of his unresolved dependency, unexpressed hostility, and other conflicting feelings, when he is being evaluated the child is placed into an uncertain situation which he perceives as threatening. Further, the authors theorized that these conflicts and the associated anxiety is heightened in a test situation in which the examiner seemingly mirrors the evaluative role-behavior of the parents. This is especially true if he fails to satisfy the child's dependency needs by withholding positive reinforcement, such as praise, encouragement and support. Their views regarding the test-anxious child are further summarized in the following quotation:
We have described the test anxious child as one who has self-depreciatory attitudes, anticipates failure in the test situation in the sense that he will not meet the standards of performance of others or himself, and experiences the situation as unpleasant -- an affective state which signifies conflict between tendencies which are conscious as well as between conscious and unconscious tendencies. We might put this in another way: in the test situation such a child is much more aware of his own covert responses than he is of the nature of the external situation which includes, of course, the stimulus task and accompanying instructions. One might say that the test anxious response has two major (and cumulative) effects: it narrows considerably the perception of the entire field and prevents a dispassionate assessment of the nature of the problem-solving task. From this way of viewing the problem it is not surprising that our initial hypothesis would be that test anxiety essentially interferes with problem-solving in the test situation (p. 20).

Sarason and his associates provided empirical support for their hypotheses. While these findings have important ramifications for personality theory, child development, and for related disciplines, in the present context, what is especially pertinent to this investigation is the hypothesis that test anxiety will interfere with performance, especially under negatively reinforcing conditions in which the examiner emphasizes the evaluative aspects of the test situation. This view is widely shared in publications cited above and below. However, the specific ways in which test anxiety interferes with intellectual test performance needs to be examined more closely. This problem will be taken up somewhat more extensively in the next section of this review.

Test Anxiety and Intelligence Test Performance

In recent years, a large number of studies have investigated the effects of anxiety (either general or specific) upon cognitive functioning. Because of difficulties stated earlier, generalizations have been difficult to establish. In addition to test performance, dependent variables which have been studied include such diverse measures as grade-point average, problem-solving ability, and degree of verbal productivity. Research of this nature has been extensively reviewed in publications cited above. Of special interest,
however, are those studies which have dealt exclusively with the relationship between anxiety and intelligence test performance.

Sarason, et al. (1960) reviewed the literature prior to 1960. These authors conclude that in spite of the lack of systematic research, the majority of the earlier studies support the general hypothesis that anxiety in some way interferes with intellectual test performance. However, they also refer to those studies in which the anticipated relationships between various anxiety measures and intelligence test scores were not obtained. In addition to methodological weaknesses, Sarason, and his colleagues felt that conflicting findings were sometimes obtained because of the failure to employ a specific measure of predispositional anxiety peculiar to test-like situations.

In their own research, these investigators employed both the General Anxiety Scale for Children (GASC) and the Test Anxiety Scale for Children (TASC). These questionnaires, which had been developed some years earlier at Yale University, were shown to be reliable and valid measures both on the basis of their own research (S. B. Sarason, Davidson, Lighthall & Waite, 1958; S. B. Sarason, et al., 1960) and that of other investigators (McReynolds, 1968; Phillips, et al., 1970; Ruebush, 1963). While scores on the two instruments tend to be positively correlated to some extent, the TASC in particular, has been found to be especially effective for measuring the proneness in children to be anxious in test-like situations.

In terms of their own research and that of other investigators, Sarason and his associates found that test anxiety did result in decrements in cognitive performance under conditions described earlier. This was not only true of intelligence test performance but also with respect to other cognitive tasks. Thus, in addition to presenting a comprehensive theory of test anxiety from which researchable hypotheses were derived, Sarason and his associates have contributed to research methodology with the development of the TASC.
and GASC. While Sarason (1966) later found it necessary to attach some cautions to conclusions drawn from research of this nature, many of his views have been widely supported.

Other writers (viz., Kirkland, 1971; Phillips, et al., 1970; Ruebush, 1963) have reviewed the literature on the effects of anxiety upon intelligence test performance. These reviews include studies both prior to and subsequent to the research of Sarason, et al. (1960). These authors conclude that the relationship between anxiety and intelligence test performance is a negative one for the most part.

Most previous studies have compared total score differences between high anxious and low anxious groups on either individual or group intelligence tests. Significant effects have been found more often when children were used as Ss, as contrasted with adults or college students (Dunn, 1968; Spielberger, 1966a; Walker, Neilsen, & Nicolay, 1965; Walker & Spence, 1964). However, research designs have varied in their sophistication and extensive-ness. Investigators have given varying attention to age, sex, social class, and other relevant variables. While some studies have employed experimental procedures (Feldman & Sullivan, 1971; Phillips, 1971; Ruebush, 1960; I. G. Sarason & Minard, 1962), most previous investigations would be classified as assessment studies. In general, low negative correlations have been obtained between total intelligence test scores and anxiety measures (e.g., Alpert & Haber, 1960; I. G. Sarason, 1963; Zweibelson, 1956). However, the obtained correlations have not always been significant or in the expected direction, especially where measures of general or manifest anxiety rather than test anxiety were used (e.g., Feldhusen & Klausmeier, 1962; Hafner, Pollie and Wapner, 1960; Rowley & Stone, 1963). Yet, according to at least one recent reviewer (Kirkland, 1971), these correlational studies, as well as studies employing other methodologies, have failed to resolve all issues involved
in anxiety research.

One important unresolved issue is the matter of cause-effect relationships. On the one hand, Klausmeier and Goodwin (1966), view the child's intellectual limitations as basic to his anxiety level. This is reflected in the following comments:

Why is it that children of lower mental ability generally demonstrate higher anxiety? High mental ability may make it possible for the child to assess the environment more realistically -- the real and present dangers in any current threatening object, situation, or person. His fears thus may be specific and identifiable. Children of lower I.Q.'s may show greater anxiety, that is, generalized fear of a nonspecific nature, because of their limited ability to evaluate the threat in objects, situations, or people that interact with them (p. 393).

This view is in contrast to the position of Sarason, et al. (1960). As stated earlier, these and other investigators present some evidence for the opposite conclusion - that anxiety is the etiologically significant factor. Hence, high anxious children may appear less bright on assessment devices only because their performance may be impaired by anxiety. However, this conclusion needs to be verified by additional research.

Phillips, et al. (1970) also comment upon this dilemma as follows:

The relationship between anxiety and intelligence is particularly important because of its causal implications. Does this substantial negative relationship (especially for children) indicate that those who are intelligent are more capable of coping with their environment and are, therefore, less anxious? Does this relationship indicate that anxious persons have a greater difficulty attending to and retaining information, specifically information that is incidentally learned? Since intelligence tests are to some extent achievement tests and are often loaded with information that would have been incidentally learned, anxious persons may be penalized by these tests. Or, does anxiety interfere with intelligence test taking per se; that is, does anxiety lower performance on the tests that would have been higher if the anxiety had not been present? These alternative explanations lead to widely divergent possibilities for remediation and prevention, but little hard evidence is presently available for distinction between them (pp. 29-30).

But in order to resolve the dilemma, it is necessary to determine whether or not test performance "would have been higher" but for the interfering effects
of anxiety. Rather than focusing upon total score differences, perhaps the consistency hypotheses, as stated in Chapter I, might shed some light upon this issue.

Consistency as the Dependent Variable

In most of the studies cited above, which have investigated the relationship between the test anxiety variable and level of test performance, total test-score differences have been used as the dependent variable. As stated earlier, little attention has been given to other aspects of test performance, e.g., consistency in item-to-item responding. The matter is an important one. First, we need some way of determining whether or not high anxious individuals who score low on the test might really be capable of scoring higher. Second, if we can assume that anxiety does cause the performance decrement, as is suggested by the trend of research referred to immediately above, then we need to know in what specific ways performance is affected in the test situation. For example, in view of previous research, do the disruptive effects of anxiety tend to be pervasive and general? Or do they tend to fluctuate and to be irregular, specific, and greater at some times more than others. Third, it is necessary to take into account the relative difficulty and content of the test items. Does the high anxious individual obtain lower scores because he attains success only on the easier items, which pose only a minimal threat; or does he also tend to fail easy items which he might otherwise be capable of solving? Does the specific content of the items influence the anxiety reaction? Finally, if the effects of anxiety tend to be irregular rather than pervasive, does this result in a greater degree of inconsistency or erraticism in item-to-item responding?

Available research provides only partial answers to these questions.
Sarason, et al. (1960) note the tendency of the high anxious child "to reflect in his responses (when he is able to respond at all) illogical or irrational ways of thinking ...(p. 162)". Perhaps this irrationality might be reflected in an erratic test response pattern. Mandler and Watson (1966) deal with the problem only incidentally in their study of the effects of anxiety on problem-solving tasks. Their views are evident in the following comments:

With some consistency low anxious subjects have been shown to perform at a higher level under conditions where success or failure is possible ... and it has been suggested ... that the inferior performance of the high anxious subjects is due to the relatively large number of task-irrelevant responses which they make. The data ... support the notion that the high anxious subject does spend more time making responses irrelevant to task solution and perhaps inimical to it (p. 278).

Mandler and Watson further suggested that the cognitive efficiency of high anxious subjects is impaired partly because of their obsessive concerns relative to the adequacy of their performance. Such individuals may tend to become so ego-involved in the task that they cannot approach the problem (or test items) with sufficient objectivity and detachment necessary to achieve consistent success. Hence, their performance may appear spotty or inconsistent. Some additional support for this notion was obtained in a study dealing with the effects of anxiety upon problem-solving behavior. Blatt (1963) found that even on an easy item low anxious Ss were more efficient than high anxious Ss. In addition, the performance effects of anxiety appeared to be transitory. However, the question of whether or not such effects would be observed in an actual test situation was not dealt with in this study.

Test Anxiety, Consistency, and Scatter

While the relationship between anxiety, cognitive efficiency, and consistency in test performance has not been widely investigated through empirical
research, it has been given considerable emphasis by clinicians. In clinical studies in the field of psychopathology, inconsistency is usually dealt with under the rubric of "test scatter." Though definitive evidence is lacking, it is widely held that consistency in performance is a reflection of intellectual efficiency associated with emotional stability. On the other hand, the degree of scatter or inconsistency is seen as being related to the degree of cognitive dysfunction associated with personality disturbances in which anxiety is a major component. This view is emphasized in two influential publications cited earlier (Rapaport, et al., 1945; Wechsler, 1958).

Partly because of their popularity and convenient breakdown into individual verbal and performance subtests, many of the conclusions about scatter have been based upon research using tests devised by Wechsler, viz., the Wechsler-Bellevue, the Wechsler Adult Intelligence Scale (WAIS), and the Wechsler Intelligence Scale for Children (WISC). According to previous findings, the degree of variability or scatter in S's performance may be looked upon as being an indication of the degree of psychopathology. This view is also presented by Rapaport, et al. (1945) in their widely read book on clinical application of individual psychological tests. Particular attention was given to the Wechsler-Bellevue, the predecessor of the WISC and WAIS. Two types of scatter or variability have been described in the literature concerning these three tests. In accord with the view of Wechsler (1958) and others, these two diagnostic measures have been described and defined succinctly by Horrocks (1964) as follows:

Another category of diagnostic significance is that of test variability which may be divided into inter-test and intra-test variability. Inter-test variability, also known as "scatter", is evidenced when an examinee does very well on some subtests and very poorly on others. Intra-test variability is shown when an examinee fails easy and passes hard items within the same subtest (p. 217).

As defined above, inter-test scatter has been used persistently by clin-
icians as a diagnostic differentiator, although findings on the validity of this indicator have been conflicting, confused, and disjunctive. Reviewers of the WISC have been especially critical of those studies which have attempted to contrast and compare the inter-test scatter patterns of loosely constituted diagnostic groups (e.g., schizophrenics vs. neurotics, delinquents vs. nondelinquents, etc.). Examples of especially critical reviews on the applicability of inter-test scatter may be found in recent publications by Cohen (1959), Littell (1960), Frank (1970), and Anastasi (1968). The critical points noted by these authors include the limited reliability of the individual subtests, their lack of specificity, the over-emphasis upon face validity, the difficulties in establishing diagnostic criteria and other weaknesses. Hence, it would appear that in connection with inter-test scatter clinicians have failed to heed cautions presented by investigators of scatter patterns on earlier tests, such as the Stanford-Binet (Lorr & Meister, 1941).

In contrast to the considerable attention which has centered upon WISC inter-test scatter, either pro or con, comparatively little attention has been directed to the topic of intra-test scatter. Allison, Blatt, and Zimet (1967), Blatt and Allison (1968), Horrocks (1964), Rapaport, et al. (1945), Wechsler (1958), and reviewers of the WISC and other Wechsler tests give it only brief mention. This lack of attention is somewhat puzzling in view of its widespread practical use by psycho-diagnosticians. Furthermore, intra-test scatter may provide a more researchable topic in that it avoids many of the pitfalls described above in relation to inter-test scatter. For example, the question of differential subtest reliability and specificity (Hopkins & Michael, 1961) is avoided.

Although no studies could be located in the literature which deal directly
with intra-test scatter on the WISC, a few researchers have investigated intra-test scatter on the adult versions of the Wechsler, i.e., the Wechsler-Bellevue and the WAIS (Hallenbeck, Fink, & Grossman, 1965; Holzberg & Deane, 1950; Nickols, 1963; Saxe, 1966; Watson, 1965). Guertin, Ladd, Frank, Rabin, & Hiester (1966), in their review of research with the WAIS for the period 1960-1965, suggest that some of these studies have offered promising leads for future research. However, there are many difficulties in attempting to generalize from these few studies. In addition to other methodological deficiencies, the quantitative measure of intra-test scatter has varied from study to study. Second, all of the studies cited above have been based upon loosely constituted groups compared on the basis of non-operationally defined psychiatric classifications, e.g., schizophrenics, organics, neurotics and "normals." Third, in those instances where significant differences were found, the subtest(s) in question differed from those identified in other studies. Fourth, none of the findings are directly applicable to educational and developmental studies using children drawn from normal populations. Fifth, none of the studies cited above has employed experimental procedures. In the main, prior investigations of intra-test scatter have primarily of an assessment nature. Sixth, for the most part, empirical findings have not been adequately related to measurement theory or to practical application. Finally, of great importance is the fact that none of the above investigations of intra-test scatter patterns has dealt with such theoretical constructs as test anxiety. These limitations, in part, served to motivate the present investigation, as will be explained further.

Implications for Test Theory

Important theoretical issues are raised by the problem of test scatter,
i.e., variability in test performance. Both Guilford (1954) and Glaser (1949, 1951, 1952) have commented upon some of these issues. Seen from a theoretical viewpoint, the problem of scatter has been related directly to the matter of internal consistency. The internal consistency of a test is a major factor in determining its reliability.

Eysenck (1947) was among the first to deal systematically with the relationship among scatter, internal consistency, reliability and the homogeneity of test items and subtests. He discussed the relationship between measurement error due to the unreliability of persons and error due to the unreliability of tests. Thus, the internal consistency of a test is determined not only by the degree to which the items are related to one another, but also by the degree of consistency with which each examinee responds to the items. Personality functioning was seen as an important component in determining response consistency. Eysenck concluded that for homogeneous (factorially pure) tests, in theory, "scatter score gives a minute-to-minute reliability (p. 121)", reflecting momentary fluctuations in an individual's performance.

Fiske and Rice (1955) presented a comprehensive discussion on the topic of intra-individual response variability. Though they did not directly deal with the concept of anxiety, they reviewed studies which indicated that variability in test performance is related to certain, as yet, unestablished "personality correlates (p. 243)." Further, Fiske and Rice suggested that variability could be increased with experimental stress. In accord with Eysenck (1947), they felt that in a test situation factors associated with response variability would affect the reliability of the test.

Sechrest & Jackson (1967) also deal with the theoretical implications of intra-individual response variability (or scatter) from a somewhat differ-
ent but not unrelated point of view. They deal with the problem from the standpoint of Berg's "Deviation Hypothesis." This hypothesis was developed primarily in conjunction with objective personality tests, but may have relevance for all types of measurement instruments. Briefly stated, the Deviation Hypothesis holds that certain individuals tend to be deviant in their response patterns regardless of the nature of the item content (Berg, 1959). However, this hypothesis was felt to be too general by Sechrest and Jackson (1967).

In an empirical attempt to examine the implications of the hypothesis, Sechrest and Jackson provide some empirical support for it. Two forms of a mental ability test (ACE) were administered to a group of subjects. For most individuals, scores on one form of the ACE tended to show moderately high positive correlations with scores on the parallel form, as would be generally predicted on the basis of earlier reliability studies. However, as was noted earlier, the scores of certain individuals on one form of the test tended to be unrelated to their scores on the alternate form. And when all scores were represented in a scatter plot, the scores of these individuals tended to deviate significantly from the regression line. Such individuals were referred to as correlation "outliers", because their "performance is predicted poorly by regression scores (p. 275)". When the sample was retested with a measure of deviant response tendencies (Perceptual Reaction Test), the correlation "outliers" again tended to show a more deviant response pattern than their more consistent counterparts. Similar conclusions were reached on the basis of other investigations described by the authors.

Sechrest and Jackson conclude that with respect to test scores "correlations between any two sets of variables will be attenuated (p. 275) to
the extent that deviant response tendencies exist in individuals within
the group that is tested. The writers account for the tendency of some
individuals to deviate in their response patterns by suggesting that such
individuals show a unique structuring of traits and abilities. They further
suggest that the problem "should be afforded increased attention in psycho-
metric methodology (p. 274)", since we need to know what factors would
account for such unique structuring of traits and abilities. This view is
reflected in the following comments:

Berg and his colleagues have performed a valuable service in emphasizing
the importance of studying deviant response patterns. However, the study
of such patterns should be increased in scope and complexity to take
into account the many ways in which different people may be deviant,
the role of different classes of content in eliciting deviant responses,
and particular types of noncritical deviation unique to a given psycho-
pathological group, among other things. New analytical methods for
treating data are required to do justice to the complexity of deviant
response patterns (pp. 276-277).

In the context of the present investigation, it is felt that test anxiety
might serve as a useful explanatory construct. Thus, on the basis of previous
research cited above, one might reasonably ask, "Do correlation 'outliers'
tend to deviate because their intense anxiety prevents them from responding
in a consistent manner?" Hopefully, this question will be answered within
the present investigation.

Loevinger (1967), in her exhaustive review of the literature, also deals
with the theoretical issues centering around the problem of reliability and
intra-individual variability or scatter. Scatter is related to the topic
of "homogeneity" or "reproducibility", while test-retest reliability is
related to the problem of measurement "stability." Loevinger, also discusses
the notion that on some tasks disturbed individuals may tend to be less con-
sistent in their performance than do normal individuals. This fact, if true,
would then attenuate the reliability coefficient obtained for a given test.
Hence, the obtained reliability coefficient would be lowered regardless of whether it is based upon test-retest comparisons or upon single-trial administrations.

The article by Loevinger, like others cited immediately above, is especially notable for its attempt to relate the problem of scatter or inconsistency to classical test theory. Her views are in accord with those stated earlier by Eysenck (1947). This is evident from the following comments:

Two quite different kinds of scatter have been studied: intra-test and inter-test scatter. In the case of intra-test scatter, the test items are assumed to be equivalent in function and the dispersion of successes and failures along the scale of difficulty (or its analog) represents S's tendency to inconsistency. In the case of inter-test scatter, the functions called on by several tests are not assumed to be identical. Many studies have attempted to show, with varying degrees of success, characteristic patterns of relatively high and low scores for different clinical syndromes... The Wechsler-Bellvue test has been most often studied in this regard, but other tests have also been used. The two types of scatter might easily be termed "pure scatter" (intra-functional). Both types of scatter have theoretical implications, but patterned scatter can be expected to be more fruitful theoretically. There are many possible ways of measuring scatter; theoretical implications of the various coefficients have not been fully explored (Loevinger, 1967, pp. 103-104)."

While one might disagree with her view that patterned (inter-functional) scatter is the more fruitful concept for research, one would not question the statement that there are important theoretical implications to scatter.

Problem of Individual Differences

In recent years there has been a growing dissatisfaction with psychometric theory because of its failure to take sufficient account of individual differences. Loevinger is critical of classical psychometrics and its "naively operational experimental-theoretical" orientation which fails to take adequate account of "intra-psychic factors" operating within "the behaving person (p. 119)." She indicates the need for further research within this problem area.
Other writers have also criticized educators and psychometricians for similar reasons. Anastasi (1967) has decried the overemphasis upon psycho-technology and the resulting gap between psychometric theory and psychological theory. She is especially concerned with the dangers of over-specialization in psychology and education, as reflected by the tendency to view the affective and cognitive domains as disparate. This view is reflected in the following comments:

...The separation between abilities and personality traits is artificial and the two domains need to be rejoined in interpreting an individual's test scores. It is now widely recognized that an individual's performance on an aptitude test, in school, on the job, or in any other context is significantly influenced by his achievement drive, his self-concept, his persistence and goal orientation, his value system, his freedom from handicapping emotional problems, and every other aspect of his so-called personality (p. 305).

Mayo (1965), in reviewing developments in item analysis techniques, also presents criticisms directed at "the overemphasis upon test development relative to the use of empirical criteria at the exclusion of consideration of psychological theory as a basis for the explanation of the empirical relationships (pp. 82-83)." It is further pointed out that traditional item analysis techniques reveal little about why examinees respond to the items as they do. In a more recent review on test theory, Bock & Wood, (1971) concluded that the tendency of psychometricians to ignore psychological theory has persisted in spite of attempts to broaden test theory. In view of growing criticisms of psychological testing, Kirkland (1971) views the tendency to ignore such matters as being especially unwise.

Cronbach (1967), from a similar perspective, has highlighted the need for a unified effort toward understanding human behavior. Such an effort, he feels, should ideally incorporate many diverse theoretical and applied disciplines within the behavioral sciences. It is within this broad frame of reference that the present investigation is being carried out. The
methodology employed in this effort will be presented in the next chapter.
CHAPTER III

METHOD

Selection of Subjects

As was noted in the introductory chapter and in the preceding review of the literature, test anxiety has been found to have a major impact upon children, especially during their intermediate years in school. During this important transitional period in a youngster's school experience, anxieties about scholastic achievement and about test-taking often become rather intense (Kagan & Kogan, 1970; Klausmeier & Goodwin, 1966; S. B. Sarason, et al., 1960). Mainly for these reasons, the sample of pupils employed in this investigation was selected from a population of fifth-graders. The three cooperating school districts from which Ss were drawn are similar in that they all are of moderate size, having enrollments of approximately 2000 to 3000 pupils. All three districts are located within a predominantly white, middle-class, suburban region located outside of the city of Chicago.

The TASC, a research questionnaire developed at the Yale University Psycho-Educational Clinic by S. B. Sarason, et al. (1958), was employed for initial screening and selection purposes. Consent to employ this instrument was obtained from the senior author. Previous research with the TASC shows it to be among the most effective, self-report measures of test anxiety in elementary school children (Alpert & Haber, 1960; McReynolds, 1968; Phillips et al., 1970; Ruebush, 1963; S. B. Sarason, et al., 1960).

The TASC was administered by one female and two male examiners. All three examiners were experienced in administering psychological tests.
The questionnaire was administered in October, 1971, over a two-week period, in regular classrooms, and under relatively standardized conditions. By prior arrangement, classroom teachers were not present while the TASC was administered. In each instance, the examiners noted that, by and large, students were quite cooperative in responding to the questionnaire. The examiner and examinee directions are presented along with the questionnaire items in Appendix A.

After the TASC had been administered and scored, Ss were selected from each end of the distribution of scores, i.e., whether on the basis of their scores they expressed a predisposition to be either high anxious (HA) or low anxious (LA) in test-like situations. The cutoff points were established at the upper and lower quartiles. A similar procedure had been used in studies by Ruebush (1960), Zweibelson (1956), and in other comparable studies cited by S. B. Sarason, et al., (1960).

In addition to the criteria described above for selection of Ss, other criteria included I.Q. and sex. Generally, only Ss of average level intelligence were selected on the basis of previously administered group-tests of mental ability used as part of the regular testing program by the districts. These group measures included the SRA-Primary Mental Abilities test, the Otis-Lennon, the Kuhlmann-Anderson and the STS-Educational Development Series. Ss were selected from an I.Q. range of 85 to 115.

With respect to the sex variable as it is related to anxiety in elementary school children, Phillips, et al. (1970), on the basis of their review of the literature, indicated that, "Sex differences in anxiety have been consistently obtained, with girls having higher scores ... (p. 9)". However, this view contrasts with a conclusion based upon a somewhat more recent review (Kirkland, 1971). This reviewer concluded that, "There is generally no relationship between anxiety and sex at the elementary school level ... (p. 318)."
Hence, in view of these incompatible findings, in the present investigation the sex variable was taken into account by including an equal number of males and females. Hence, the entire sample consisted of 48 (24 LA and 24 HA) boys and 48 (24 LA and 24 HA) girls.

Matching Subjects

With respect to the matching of Ss on the control variables, Ss were paired on the basis of sex, I.Q., and anxiety level, i.e., whether they were either LA or HA. One member of each pair was then randomly assigned to a given treatment group; the other member, to the alternative treatment group. In order to verify the expectation that there would be no significant differences with respect to matching of 96 Ss on the basis of C.A. and prior I.Q., two separate, three-way, analyses of variance (ANOVA) were performed using a 2 X 2 X 2 factorial design.

First, with regard to C.A., Table 1 presents ANOVA results with C.A. in months used as the criterion measure. The grand mean of the entire sample was ten years, seven months. As anticipated, at the .05 level of significance, no significant between-group differences were obtained on C.A. for sex, anxiety level, or assignment to treatment conditions. Table 1 also reveals that there were no significant interactions among factors or among categories (i.e., levels) of factors on this control variable.

Second, with respect to prior I.Q., the ANOVA results are presented in Table 2. The mean for the entire sample was 103.74 with a standard deviation of 7.49. Again, as anticipated, Table 2 reveals no significant between-group differences or interactions with respect to matching of Ss on I.Q. in terms of sex, anxiety level, or assignment to treatment groups (p. > .05).

In conclusion, the ANOVA results regarding the two control variables,
TABLE 1
ANOVA on Matching of Ss on the Basis of C.A., according to Sex, Assignment to Treatment Condition, and Anxiety Level

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (S)</td>
<td>20.167</td>
<td>1</td>
<td>20.167</td>
<td>.171</td>
</tr>
<tr>
<td>Treatment (T)</td>
<td>73.500</td>
<td>1</td>
<td>73.500</td>
<td>.625</td>
</tr>
<tr>
<td>Anxiety Level (A)</td>
<td>9.375</td>
<td>1</td>
<td>9.375</td>
<td>.071</td>
</tr>
<tr>
<td>S x T</td>
<td>240.667</td>
<td>1</td>
<td>240.667</td>
<td>2.047</td>
</tr>
<tr>
<td>S x A</td>
<td>77.042</td>
<td>1</td>
<td>77.042</td>
<td>.655</td>
</tr>
<tr>
<td>T x A</td>
<td>45.375</td>
<td>1</td>
<td>45.375</td>
<td>.386</td>
</tr>
<tr>
<td>S x T x A</td>
<td>45.360</td>
<td>1</td>
<td>45.360</td>
<td>.386</td>
</tr>
<tr>
<td>Within</td>
<td>10349.781</td>
<td>88</td>
<td>117.611</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10861.266</td>
<td>95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .05
TABLE 2
ANOVA on Matching of Ss on the Basis of
I.Q., according to Sex, Assignment
to Treatment Condition, and
Anxiety Level

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex(S)</td>
<td>.260</td>
<td>1</td>
<td>.260</td>
<td>.0046</td>
</tr>
<tr>
<td>Treatment (T)</td>
<td>.510</td>
<td>1</td>
<td>.510</td>
<td>.0091</td>
</tr>
<tr>
<td>Anxiety Level (A)</td>
<td>.094</td>
<td>1</td>
<td>.094</td>
<td>.0017</td>
</tr>
<tr>
<td>S x T</td>
<td>.010</td>
<td>1</td>
<td>.010</td>
<td>.0002</td>
</tr>
<tr>
<td>S x A</td>
<td>14.261</td>
<td>1</td>
<td>14.261</td>
<td>.254</td>
</tr>
<tr>
<td>T x A</td>
<td>1.260</td>
<td>1</td>
<td>1.260</td>
<td>.022</td>
</tr>
<tr>
<td>S x T x A</td>
<td>.003</td>
<td>1</td>
<td>.003</td>
<td>.0001</td>
</tr>
<tr>
<td>Within</td>
<td>4934.062</td>
<td>88</td>
<td>56.069</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4950.461</td>
<td>95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .05
C.A. and I.Q., offer assurance that the 96 Ss were adequately matched. Hence, any significant between-group differences on measures of the dependent variable would not likely be attributed to spurious effects associated with differences in either C.A. or I.Q.

Parenthetically, the 2 X 2 X 2 factorial design employed in verifying that there were no significant between-group differences on the control variables is similar to that employed for testing for differences on certain measures of the dependent variable. These results will be presented in Chapter IV. The three-way ANOVA design used in this study is taken from Lindquist (1953).

Dependent Variable

As had been stated earlier, the dependent variable investigated in this study was inconsistency in intellectual performance. In order to make the study manageable in terms of both practical and theoretical considerations, the Comprehension subtest of the WISC was singled out as the intellectual test most applicable to this investigation. The items comprising this subscale consist of a number of "why" or "what to do" types of questions calling for practical solutions to everyday problem situations. Factor analytic data suggest that the Comprehension subtest presents a rather good measure of general intelligence in terms of the extent to which it loads on the G factor. Also, in pre-adolescent youngsters it tends to be less a measure of learning gained through formal education than do other verbal subtests in the WISC battery (Cohen, 1959). Because the items call for on-the-spot reasoning, they are viewed as being rather sensitive to temporary disruptions in problem-solving ability, according to Glasser and Zimmerman (1967). Furthermore, success on the items tends to be reflective
of social and moral judgment (e.g., "What is the thing to do if a fellow much smaller than yourself starts to fight with you?" "Why should a promise be kept?"). Partly for these reasons Glasser and Zimmerman state that the test is "particularly vulnerable to maladjustment" and to "transient emotional reactions" and "gives knowledge of the child's coping ability (p. 53)." Along these lines, a preliminary study undertaken prior to the present investigation partially supported this view, since it had been found that HA Ss did show a significantly greater degree of inconsistency in item-to-item responding than did LA Ss. This finding was not true, however, with respect to the other WISC subtests under consideration.

Several other considerations were taken into account in selecting the Comprehension subtest. In previous studies reviewed by Glasser and Zimmerman, it had been found that with only minor exceptions the Comprehension items are quite adequately placed with respect to graded level of difficulty. This fact was not as true of some of the other WISC subtests. The ordering of the items in terms of difficulty, then, would be an important pre-condition for testing the hypotheses. While the standardization data suggests that the split-half reliability of the Comprehension subtest is slightly less than that of the other verbal subscales of the WISC (with the exception of Digit Span); nevertheless, its reliability is regarded as acceptable for an intelligence test (Quereshi, 1968; Wechsler, 1949). Furthermore, for the purpose of this investigation, a cognitive test of moderately high reliability with a measurable degree of unexplained variance would conceivably be more sensitive to the effects of nonintellectual factors than a stable measure of very high reliability.
Procedure

The treatment factor consisted of the conditions under which the individual intelligence test (Comprehension) was administered. Two levels of the treatment variable were compared, viz., nonstressful (N) vs. stressful (S) test administration.

Certain similarities in procedure were employed to minimize the influence of extraneous variables and to insure comparability between the nonstressful and stressful treatment conditions. For example, all Ss were individually tested during the first three weeks of November, 1971, or about three-to-six weeks after the TASC had been administered. Testing was done by a male examiner with whom Ss were previously unacquainted. Each S was tested for about fifteen minutes in his home school building. Ss were seen in more or less random order in accordance with their respective classroom schedules. To eliminate possible experimenter bias, E tested S "blind", i.e., without knowledge of his previous group I.Q. score, TASC score, or anxiety level as recommended by a number of investigators (Frank, 1970; Littell, 1960; Masing, 1968; Sattler & Theye, 1967; Sattler, Winget, & Roth, 1969). Since it was often necessary to use whatever rooms were available at a given time, conditions varied somewhat. However, this lack of perfect control is viewed as a usual concomitant of experimental research undertaken within a naturalistic setting. Advantages and drawbacks of research of this nature have been discussed elsewhere. However, to insure further comparability, the arrangement of desks and chairs was kept as uniform as possible. All Ss were asked to defer their questions until after the test. A special recording form was devised to insure orderly and systematic collection of the data (see Appendix A). Use of this form also allowed for full and close recording of each S's responses.
With respect to differential treatment conditions, in the nonstressful testing situation, conditions approximated a more or less typical testing situation. As recommended by almost all test manuals and basic texts on test administration, the threat of evaluative testing was diminished insofar as possible by E's attempts to establish a cooperative, friendly, nonjudgmental atmosphere. For example, Ss were addressed by their first names. A three-to-five minute conversation period preceded administration of the test in order to establish rapport. During this time, discussion centered on S's interests, achievements, and activities. After rapport had been established, the following directions were then presented:

I've asked your teacher to let you come here so we can go through a few things together. I'll be asking you some questions about a lot of different things. I'll also be seeing other kids from your room, because we want to find out how most fifth-graders answer these questions. Although the results won't go into your records, please try to give the best answers you can. O.K.? Let's begin.

In presenting these instructions, E avoided any reference to the word, test.

In spite of the generally relaxed and permissive atmosphere which was established, standardized testing procedures were closely adhered to in the manner that they are presented in the test manual (Wechsler, 1949). As is the usual procedure, Ss were given appropriate encouragement, praise, and support for their efforts. For example, at various points throughout the test, Ss were told that they were doing well, or that their responses were "good." If an item proved too difficult, a supportive comment was made, such as, "That was a hard one; you'll probably be able to get that one when you're a little older." Ambiguous responses requiring clarification were questioned in the neutral, nonsuggestive manner recommended in the manual. For example, E made such comments as, "Tell me more about that," or "Please explain further." However, careful effort was made so as to avoid prompting responses or to avoid rejecting inferior responses.
With respect to the stressful treatment condition, a number of experimental studies have created stressful test conditions, e.g., by presenting ego-involving instructions, by deriding Ss, by threatening punishment (e.g., school failure), and by withholding all feedback, support, and encouragement (Dunn, 1968; Egeland, 1967; Fiske & Rice, 1955; Phillips, et al., 1970; Phillips, 1971; I. G. Sarason & Minard, 1963; S. B. Sarason, et al., 1960; Walker, Nielsen, & Nicolay, 1965; Walker, Sannito, & Firetto, 1970; Walker & Spence, 1964). A combination of these methods was used in the present study to establish a stressful test situation.

Under the stressful treatment, S was greeted coldly and impersonally by E and immediately presented with the following instructions:

Sit down please. Your name? I'm here to give you a test. This test is used for grade placement levels for the school year. The test is important because it helps us to find out how intelligent a boy or girl is, as well as how well he (she) can learn. Some children score higher than others because they're smarter or can learn better. I don't know how well you will do on the test, but it's important for you to do the very best that you can. Do you understand? Let's begin with these questions.

The test items were then presented in a cool, aloof manner but in the order and fashion prescribed in the WISC manual (Wechsler, 1949). While all praise or encouragement was withheld, each S was given sufficient time to formulate and to clarify his responses. As in the nonstressful situation, ambiguous responses were questioned in the neutral fashion suggested in the manual; leading questions were avoided.

Upon completion of testing, whether stressed or unstressed, each S was informed that he had done well regardless of his actual performance. As might be anticipated, those Ss who took the test under stressful conditions were especially relieved when informed that the results would not be included in their records. Each S was then questioned briefly about his reactions to the test. He was subsequently invited to raise any pertinent questions that
he might have had. These questions were answered frankly but in such a way as to avoid revealing the purpose of the experiment. S was also asked not to inform his classmates of what had taken place. There were no known instances in which any of the Ss had prior knowledge of the purpose of the experiment.

Scoring of the test responses was delayed until after the data had been collected. To further minimize the possibility of experimenter bias, a "blind" scoring procedure was employed in addition to "blind" testing. All responses were scored closely in accordance with the examples provided in the test manual (Wechsler, 1949). As recommended by Cronbach (1970), a supplementary manual (Massey, 1967) was employed to assist in scoring of ambiguous responses not clarified after questioning. The obtained data were then subjected to statistical analysis.

Summary

The intent of this experimental investigation was to compare the effects of test anxiety and test conditions upon consistency in intelligence test performance. The sample employed in this investigation consisted of 96 pupils from several public, elementary schools, located within a predominantly white, middle-class, suburban region. Ss were paired on the basis of sex, age, grade, group I.Q. score, and test anxiety level as measured by the TASC, i.e., whether they were either LA or HA. All Ss were individually tested with the Comprehension subtest of the WISC under either of two alternative test conditions. Of the 48 LA Ss, half were randomly assigned to a nonstressful treatment condition. A similar procedure was followed with the remaining 48 HA Ss. A "blind" testing and scoring procedure was employed.

Ss were compared on several measures with respect to their level of performance and consistency in item-to-item responding. The results of these statistical comparisons will be presented in Chapter IV. Finally, the im-
Applications of these findings will be presented in Chapter V.
CHAPTER IV

RESULTS

The present study was designed to examine the effects of test anxiety and test conditions upon consistency in intelligence test performance. Several hypotheses were presented in Chapter I. The predictions were as follows: (1) Regardless of the conditions under which they are tested, in general, HA Ss will perform more inconsistently than LA Ss. (2) Regardless of anxiety level, in general, Ss tested under stressful conditions will perform more inconsistently than Ss tested under nonstressful conditions. (3) Finally, there will be a significant interaction effect between an S's predisposition to be anxious (anxiety level) and the conditions under which he is tested (treatment effect). Accordingly, the highest degree of inconsistency will be found in the HA subgroup tested under stressful conditions; whereas, the lowest degree will be found in the LA subgroup tested under nonstressful conditions.

Two methodological strategies were employed in testing the experimental hypotheses. The first strategy made use of correlational analysis. Groups were compared on conventional, psychometric measures of internal-consistency reliability. The major objective of this use of single-trial reliability was to initially determine whether or not the effects of the independent variables might be evident from group differences on an overall, standardized, inter-individual measure of consistency in item-to-item responding. Accordingly, it was assumed that the differential effects attributed to the treatment con-
ditions and to anxiety level would be manifested in significant differences among the resulting, internal-consistency, reliability coefficients obtained for each group and for each subgroup.

The second strategy for testing the hypotheses employed two related measures of intra-test scatter (which was defined earlier) as criterion scores in two, separate, analyses of variance. The major objective of the second method was to determine whether or not the effects of the independent variables would be evident from specific, clinically-derived, intra-individual measures of consistency (or, rather, inconsistency) in item-to-item responding. In addition to serving as an additional method of testing the experimental hypotheses, this method was also used to subject a commonly used diagnostic measure to experimental evaluation. Accordingly, it was assumed that this approach would also serve to test the validity and utility of intra-test scatter as a workable concept in the field of educational and psychological measurement, from both a pragmatic and a theoretical standpoint.

The data, including those presented in Chapter III regarding ANOVA on the control variables, C.A. and I.Q., were analyzed on an IBM 360 computer. The results obtained using the two methodological strategies described above will be presented separately in later sections of this chapter. The implications of these findings will be discussed in detail in the following chapter.

ANOVA on the Effects of the Independent Variables Upon Level of Performance

Prior to applying the statistical tests of the experimental hypotheses pertaining to consistency in performance, initial comparisons were made primarily in order to determine whether or not differences in consistency measures
could arise independently of differences in total score or level of performance. Another reason for comparing Ss upon differences in total score was to determine whether or not the results of the present study were in agreement with those previously cited investigations on test anxiety in which differences in level were found.

With respect to the actual procedures employed in measuring level of performance, the total raw-score was derived in the usual manner for the Comprehension test. It consisted of the sum of the number of points for both fully credited (two-point) items and partially credited (one-point) items. Each S's raw score was then converted to its normative, age-scaled score, having a mean in the population of 10 and a standard deviation of 3. The derived scale-score was then used as the criterion measure in a 2 X 2 X 2 factorial design. The mean scale score on the Comprehension test for the entire sample of 96 Ss was 10.91 with a standard deviation of 3.24.

The results of the three-way ANOVA are presented in Table 3. None of the F ratios are significant at the .05 level. There were no significant, between-group differences associated with sex, treatment, or anxiety level. There were also no significant first-order or second-order interactions among the three factors or among levels of factors, respectively. Thus, these obtained nonsignificant differences in total score on the intellectual measure fail to reveal any interfering or facilitating effects associated with test anxiety level or with test conditions that would be evident from S's level of performance. On this basis, it was assumed that any predicted group differences that might have occurred on the consistency indices arose independently of level of performance.
TABLE 3
ANOVA: Differences in Total Score Associated
With Sex, Test Conditions, and Anxiety Level

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (S)</td>
<td>17.510</td>
<td>1</td>
<td>17.510</td>
<td>1.543</td>
</tr>
<tr>
<td>Test Conditions (T)</td>
<td>.510</td>
<td>1</td>
<td>.510</td>
<td>.045</td>
</tr>
<tr>
<td>Anxiety Level (A)</td>
<td>3.760</td>
<td>1</td>
<td>3.760</td>
<td>.331</td>
</tr>
<tr>
<td>S x T</td>
<td>3.010</td>
<td>1</td>
<td>3.010</td>
<td>.265</td>
</tr>
<tr>
<td>S x A</td>
<td>3.010</td>
<td>1</td>
<td>3.010</td>
<td>.265</td>
</tr>
<tr>
<td>T x A</td>
<td>14.260</td>
<td>1</td>
<td>14.260</td>
<td>1.257</td>
</tr>
<tr>
<td>S x T x A</td>
<td>17.516</td>
<td>1</td>
<td>17.516</td>
<td>1.544</td>
</tr>
<tr>
<td>Within</td>
<td>998.578</td>
<td>88</td>
<td>11.347</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1058.156</td>
<td>95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
The Effects of the Independent Variables
Upon Internal Consistency Reliability

Reliability measures obtained on the basis of a single test administration are referred to as coefficients of internal consistency. Usually these reliability measures are used in test standardization to provide information on the test's homogeneity, i.e., how consistently the items measure a single trait or group of traits when administered to a sample of more or less "typical" examinees. However, Cureton (1967), among others, indicates that reliability measures are influenced about as much by the individual reactions of the examinees to the items as by the actual content and nature of the items. Thorndike (1951), Nunnally (1967), Guilford (1954), and others have presented comprehensive discussions on subject variables influencing an examinee's reactions to the items. These individual differences among testees thus greatly influence both the intercorrelations among items and the resulting reliability measurements for the test as a whole. However, such measures are typically used to provide information only about the test, though potentially they could also provide a great deal of information about the examinees and their reactions to the test. This would especially be the case when there are a priori reasons for anticipating group differences as in a study by Webb and DeHann (1951) and in the present experimental investigation. White and Saltz (1967), in their article on the measurement of reproducibility, a topic closely associated with internal consistency, suggest this approach in the following quotation:

There is no reason why the techniques of computing reproducibility or single-trial reliability cannot be reversed to yield coefficients about the homogeneity of subjects, instead of test items. It is surprising that this has not been done more often... Lack of reproducibility in a response matrix is just as likely to be due to heterogeneity in the population tested, as to heterogeneity in the test items (p. 255).
In the present investigation, the lead of White and Saltz was followed in that internal consistency data was used to focus upon the examinees, rather than solely upon the test. But, while White and Saltz state that measures of internal consistency and reproducibility (e.g., those of Guttman or Loevinger) provide similar results, other authors have disagreed with this position. Both Nunnally (1967, p. 66) and Thorndike (1951) present arguments against the use of reproducibility measures and in favor of the use of more firmly grounded, internal-consistency, reliability measures.

There appears to be some lack of agreement among measurement specialists as to the measure of internal-consistency reliability most appropriate for a test such as Comprehension. While a split-half, reliability measure was used in the initial standardization of the WISC battery (Wechsler, 1949) and in at least one follow-up study (Quereshi, 1968), this technique is said to have definite limitations over other procedures. According to Guilford (1954), a given split-half procedure for measuring reliability (obtained, e.g., by dividing the test into odd-numbered versus even-numbered items) provides only one of "many ways of splitting n items into two sets of n/2 items each (p. 380)." Because of this limitation, Guilford states that an alternative method was developed which provides an average of all possible split-half coefficients. This method involves the use of the well known KR-20 formula. Incidentally, similar measures providing results comparable with those obtained using KR-20 have been proposed by Hoyt (1941) and by Cronbach and Azuma (1962). Nunnally (1967), taking an extreme position in favor of the use of measures such as KR-20, states, "The only reason for employing a split-half method occurs when the items are scored not dichotomously, but on three or more points (p. 214)." Hence, with few exceptions, Nunnally definitely favors the use of the Kuder-Richardson procedure. On the other hand, Cuerton (1967) favors a broader use of split-half reliability
in that this technique requires less rigid assumptions than its more sophisticated counterpart. He concludes that the choice of methods should depend largely on the degree of homogeneity and the factorial composition among the test items. Of course, this widely held position does not take into account the notion that homogeneity and factorial composition may vary greatly with the sample tested -- an assumption which is basic to the present investigation. Hence, mainly in view of these seemingly divergent opinions, both techniques were employed simultaneously in the present study to compare Ss on the degree of consistency in item-to-item responding. In addition, the use of both reliability measures allows comparison of the different measurement properties of both techniques.

The results obtained using these two reliability measures are presented below. The use of both techniques required rescoring of item responses on a dichotomized, pass-fail basis. Where the split-half method was employed, the test was subdivided into odd-numbered items versus even-numbered items, as is the usual procedure. Both the uncorrected Pearson product-moment correlation coefficients between scores on the odd-versus even-numbered items and the corrections for test length obtained using the Spearman-Brown formula are presented in the tables. These values are presented along with findings obtained using the KR-20 formula for the purpose of comparison. All tests of significance, however, were based upon the uncorrected Pearson r's for the half-length tests, much as had been done in a related study (Webb & DeHann, 1957). In each case, two tests of significance were employed. In brief, the first test was based upon the null hypothesis that each of the obtained Pearson r's represents an estimate of the hypothetical population value (rho), which is assumed to be equal to zero (H₀: rho = 0). The second test of significance was made with respect to the differences between the obtained Pearson r's under the null hypothesis that both values represent
estimates of the same population value \( H_0: r_1 - r_2 = 0 \). These two tests were made by converting the obtained \( r \)'s to Fisher \( z \)' scores and by testing the null hypotheses in the manner described in most standard statistical texts (e.g., Edwards, 1967). All tests were made at or below the .05 level using two-tailed tests of significance.

**Sex Differences**

Prior to testing the hypotheses using the correlational methods, for reasons outlined earlier, it was deemed necessary to analyze the data with regard to possible sex differences. The correlational data pertaining to sex differences are presented in Table 4. The Pearson \( r \)'s for the correlations between Ss' scores on the odd- versus even-numbered items are presented along with the corrections for test length using the Spearman-Brown formula. Under the null hypotheses that \( \rho = 0 \), both Pearson \( r \)'s are significant at or below the .05 level \( (df_1 = df_2 = 46) \). But while the split-half, reliability coefficient is somewhat higher for girls than it is for boys, the discrepancy was not found to be statistically significant at the .05 level \( (df_1 = df_2 = 45; z = 78; p < .44) \). The discrepancy is even less marked for the KR-20 \( r \)'s both of which compare favorably with SB \( r \)'s! Hence, there is no evidence of any significant sex differences with respect to present measures of consistency in item-to-item responding. On this basis the sex variable may be disregarded with respect to formal testing of the hypotheses.

**Test Anxiety and Test Conditions**

With respect to hypothesis one, it had been predicted that HA Ss, regardless of the conditions under which they had been tested, would perform more inconsistently than LA Ss. The correlational data pertaining to this hypothesis are presented in Table 5. Inspection of the table indicates that all
TABLE 4

Sex Differences in Internal Consistency as Measured by the Correlation Between the Odd- and Even-Numbered Items (r_{ppm}), and the Spearman-Brown (r_{SB}) and Kuder-Richardson (r_{KR20}) Formulas

<table>
<thead>
<tr>
<th>Sex</th>
<th>n</th>
<th>r_{ppm}</th>
<th>r_{SB}</th>
<th>r_{KR20}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>48</td>
<td>.308*</td>
<td>.470</td>
<td>.545</td>
</tr>
<tr>
<td>Girls</td>
<td>48</td>
<td>.448**</td>
<td>.619</td>
<td>.575</td>
</tr>
</tbody>
</table>

* p < .05
** p < .01
### TABLE 5

Differences in Internal Consistency Associated with Anxiety Level as Measured by the Correlation Between the Odd- and Even-Numbered Items (r<sub>ppm</sub>), and the Spearman-Brown (r<sub>SB</sub>) and Kuder-Richardson (r<sub>KR20</sub>) Formulas

<table>
<thead>
<tr>
<th>Anxiety Level</th>
<th>n</th>
<th>r&lt;sub&gt;ppm&lt;/sub&gt;</th>
<th>r&lt;sub&gt;SB&lt;/sub&gt;</th>
<th>r&lt;sub&gt;KR20&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA</td>
<td>48</td>
<td>.521**</td>
<td>.685</td>
<td>.665</td>
</tr>
<tr>
<td>HA</td>
<td>48</td>
<td>.221</td>
<td>.362</td>
<td>.446</td>
</tr>
</tbody>
</table>

**p < .01
of the obtained internal-consistency coefficients are in the predicted directions. Using a two-tailed test \((df_1 = df_2 = 46)\), the uncorrected Pearson \(r\) obtained for the LA group is significantly higher than the hypothetical population value of zero \((p < .01)\). On the other hand, the Pearson \(r\) obtained for the HA group is not significantly different from zero. In addition, it is evident that the KR-20 \(r\)'s for the test as a whole compare favorably with the values obtained using the Spearman-Brown correction formula. However, while these findings are as predicted, the discrepancy between the two product-moment (odd-even) correlation coefficients is not substantially large enough to be regarded as significant at the .05 level \((df_1 = df_2 = 45; z = 1.67; p < .10)\). Hence, the present correlational findings are only partially supportive of the first experimental hypothesis.

The correlational data pertaining to hypothesis two, by which differences in consistency associated with test conditions had been predicted, are presented in Table 6. Interestingly enough, the data resemble those presented above with respect to the anxiety variable. Again, using a two-tailed test of significance with 46 \(df\) in both groups, the Pearson \(r\) obtained for the group tested under nonstressful conditions \((N)\) is significantly different from zero \((p < .01)\). In contrast, the \(r\) obtained for the group tested under stressful \((S)\) conditions is not significantly different from zero even at the .05 level. As was the case earlier, the KR-20 \(r\)'s are again in the predicted directions. However, while these findings are as anticipated, again the discrepancy between the Pearson \(r\)'s was not found to be sufficiently large to be regarded as statistically significant at the .05 level with a two-tailed test \((df_1 = df_2 = 45; z = 1.61; p < .11)\). On this basis, the correlational data fail to provide clear-cut support for the second hypothesis, though it would not be regarded as untenable.
TABLE 6

Differences in Internal Consistency Associated with Test Conditions as Measured by the Correlation Between the Odd- and Even-Numbered Items ($r_{ppm}$) and the Spearman-Brown ($r_{SB}$) and Kuder-Richardson ($r_{KR20}$) Formulas

<table>
<thead>
<tr>
<th>Test Conditions</th>
<th>n</th>
<th>$r_{ppm}$</th>
<th>$r_{SB}$</th>
<th>$r_{KR20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonstress (N)</td>
<td>48</td>
<td>.517**</td>
<td>.682</td>
<td>.631</td>
</tr>
<tr>
<td>Stress (S)</td>
<td>48</td>
<td>.229</td>
<td>.373</td>
<td>.484</td>
</tr>
</tbody>
</table>

** $p < .01$
The third hypothesis had predicted an interaction effect between anxiety level and test conditions. Accordingly, it had been anticipated that the highest degree of internal consistency reliability would be found in the LA-N (low anxiety, nonstressed) subgroup, while the lowest $r$ (least consistency) would be found in the HA-S (high anxiety, stressed) subgroup. Such an interaction effect might be inferred through inspection of the correlational data presented in Table 7. As would be anticipated, the Pearson $r$'s obtained for both LA subgroups are significantly higher than the hypothetical population value of zero under $H_0$. Also as anticipated, the $r$'s obtained for both HA subgroups do not reach a statistically significant level. Once again, the $r$'s obtained using KR-20 are also in the predicted directions. These findings, then, would provide some support for the hypothesis. Parenthetically, it is also worth noting that of all the Spearman-Brown $r$'s obtained thus far only the value of .73 obtained for LA-N subgroup reaches the $SB$ $r$ obtained in the standardization sample by Wechsler (1949). However, although the discrepancy between the LA-N and the HA-S subgroup was most pronounced and in the direction predicted by the hypothesis, under $H_0$: $r_1 - r_2 = 0$, this discrepancy would not be viewed as statistically significant ($df_1 = df_2 = 21; z = 1.82, p < .07$). And, while the discrepancies between the $r$'s obtained for the LA-N and LA-S subgroups and for the HA-N and HA-S subgroups are in the anticipated directions, these discrepancies are also not significant at the .05 level, using the above tests.
TABLE 7
Differences in Internal Consistency Associated with the Interaction Between Test Anxiety and Test Conditions as Measured by the Correlation Between the Odd- and Even-Numbered Items \( (r_{ppm}) \) and the Spearman-Brown \( (r_{SB}) \) and Kuder-Richardson \( (r_{KR20}) \) Formulas

<table>
<thead>
<tr>
<th>Anxiety Level-Treatment Condition</th>
<th>n</th>
<th>( r_{ppm} )</th>
<th>( r_{SB} )</th>
<th>( r_{KR20} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA-N</td>
<td>24</td>
<td>.585**</td>
<td>.738</td>
<td>.683</td>
</tr>
<tr>
<td>LA-S</td>
<td>24</td>
<td>.415*</td>
<td>.586</td>
<td>.606</td>
</tr>
<tr>
<td>HA-N</td>
<td>24</td>
<td>.381</td>
<td>.551</td>
<td>.545</td>
</tr>
<tr>
<td>HA-S</td>
<td>24</td>
<td>.108</td>
<td>.196</td>
<td>.387</td>
</tr>
</tbody>
</table>

* \( p < .05 \)

** \( p < .01 \)
Summary of Findings with Respect to Internal Consistency

In summary, the correlational data provide only partial support for the hypotheses. While the findings were as predicted under $H_0: r_{ppm} = \rho = 0$, under $H_0: r_1 - r_2 = 0$, the hypotheses could not be verified at conventional levels of significance with the present sample. First, with respect to the test anxiety variable (Hypothesis One), as predicted the Pearson $r$ obtained for the LA group reached a level which is significantly higher than the hypothetical $\rho$ value of zero under $H_0$. However, this was not true of the Pearson $r$ obtained for the HA group. Although these findings are as predicted, the difference between these two $r$'s was not sufficiently large as to be regarded as significant at the conventional .05 level using a nondirectional (two-tailed) test of the hypothesis. Second, with respect to test conditions (Hypothesis Two), as predicted the $r$ obtained for the group tested under stress did not reach an acceptable level of significance; whereas, the $r$ obtained for the nonstressed group did. However, again the discrepancy between the two internal consistency measures was not found to be sufficiently large enough to wholly support the second hypothesis. A similar situation prevailed with respect to subgroup differences, indicating the extent of the interaction between test anxiety level and test conditions (Hypothesis Three). The discrepancies between subgroups LA-N and LA-S and between subgroups HA-N and HA-S are in the predicted directions. Of even greater import is the fact that the Pearson $r$ for the LA-N subgroup is significant, while the $r$ obtained for the HA-S subgroup is not. Although the discrepancy between the two $r$'s is sizable, this discrepancy is not statistically significant. With respect to differences in internal consistency as measured by KR-20, the findings are for the most part consistent with findings obtained using the odd-even correlation coefficients. There was no
evidence of any significant sex differences in consistency on the basis of the correlational data.

The Effects of the Independent Variables

Upon Intra-test Scatter

The first part of this investigation revealed that there were differences in the predicted direction among Ss on internal-consistency, reliability measures. These inter-individual differences were attributed to the effects of the independent variables, though the experimental hypotheses were only partially supported in view of the fact that differences in the correlation coefficients were not always significant. The second strategy for testing the hypotheses made use of clinically derived, intra-individual measures of consistency in item-to-item responding (intra-test scatter). Intra-test scatter was described earlier in relation to ability tests consisting of items graded in difficulty. The degree of intra-test scatter is determined by S's tendency to fail easy items while he passes difficult items. Actually, two separate but related measures of intra-test scatter were developed. Both measures took into account the number of 'runs' in S's item-to-item performance.

Dichotomized Scoring

With respect to the first measure, according to Siegel's (1956) definition, a run may be regarded as either a set of consecutive correct responses or a set of consecutive incorrect responses. This usage, of course, assumes dichotomous scoring of items on a pass-fail basis, as was true when the internal consistency measures were employed. Parenthetically, Watson (1965) also used runs in this way as a criterion measure in his clinical study of intra-test scatter involving adult psychiatric patients. Hence, with this
initial measure of intra-test scatter, all responses which would otherwise be only partially credited (one-point responses) were scored as 'pass.' Accordingly, only fluctuations in S's item-to-item responding from pass-to-fail or from fail-to-pass were taken into account by the present measure of intra-test scatter.

As had been done to test for significant differences in total score, a three-way (2 X 2 X 2) ANOVA was the statistical test employed to test for significant between-group differences on intra-test scatter. The present criterion measure consisted of each S's total number of runs based upon dichotomous scoring of responses. It was assumed that the more inconsistent S's performance, the greater the number of fluctuations in performance, the greater the number of runs, and the greater the amount of intra-test scatter. The mean number of runs for the entire sample was 5.20 (S.D. = 1.94). The results of ANOVA are presented in Table 8. For this measure of intra-test scatter, no significant between-group differences were obtained with respect to sex, treatment, and anxiety level (p > .05). There were also no significant interactions among these three factors or among levels of factors. In this respect the findings are consistent with the results presented earlier with respect to comparisons on level of performance and when total score was used as the criterion measure. Hence, the results obtained using the present measure of intra-test scatter fail to support the experimental hypotheses.

Standardized Scoring

As an additional test of the hypotheses, another criterion measure of intra-test scatter was generated from the raw data. This additional measure was similar to the above, except that instead of dichotomized scoring of items, partially credited responses were taken into account. In other words,
TABLE 8

ANOVA: Differences in Intra-test Scatter Associated with Sex, Test Conditions, and Anxiety Level with Runs Used as the Criterion Measure (Responses Scored Pass or Fail)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (S)</td>
<td>.094</td>
<td>1</td>
<td>.094</td>
<td>.023</td>
</tr>
<tr>
<td>Test Conditions (T)</td>
<td>.010</td>
<td>1</td>
<td>.010</td>
<td>.003</td>
</tr>
<tr>
<td>Anxiety Level (A)</td>
<td>1.260</td>
<td>1</td>
<td>1.260</td>
<td>.309</td>
</tr>
<tr>
<td>S x T</td>
<td>5.510</td>
<td>1</td>
<td>5.510</td>
<td>1.350</td>
</tr>
<tr>
<td>S x A</td>
<td>1.260</td>
<td>1</td>
<td>1.260</td>
<td>.309</td>
</tr>
<tr>
<td>T x A</td>
<td>3.760</td>
<td>1</td>
<td>3.760</td>
<td>.922</td>
</tr>
<tr>
<td>S x T x A</td>
<td>.091</td>
<td>1</td>
<td>.091</td>
<td>.023</td>
</tr>
<tr>
<td>Within</td>
<td>359.249</td>
<td>88</td>
<td>4.082</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>371.236</td>
<td>95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .05
a run was redefined either as a consecutive set of responses on which full credit (two-points) was attained, or a consecutive set on which partial credit (one-point) was attained, or a set on which no credit (zero-points) was attained. Again, it follows that in his performance, the greater the number of runs (as defined in the present manner), the greater the amount of intra-test scatter or inconsistency in S's item-to-item responding. However, instead of basing intra-test scatter upon fluctuations in performance between passed and failed items, using the present criterion measure, one assumes that any instance in which S fails to obtain full credit will be viewed as a significant fluctuation. Hence, if S fails to attain full credit for an easy item, this failure would then contribute to the degree of variability or intra-test scatter in his overall item-to-item responding. Accordingly, it was felt that the present measure would prove to be more sensitive to fluctuations in performance.

The results obtained using a three-way ANOVA on the above criterion measure are presented in Table 9. The grand mean for the entire sample was 7.33 with a standard deviation of 1.84. Once again, at the .05 level of significance no significant main effects or their interactions are seen in terms of the present criterion measure.

Conclusions Regarding Intra-test Scatter

With respect to the implications of the above findings, in contrast to the internal consistency data, present measures of intra-test scatter fail to support the notion that variability in test performance is associated with either test conditions (induced stress) or with the predisposition to be anxious in test-like situations. This conclusion would appear to hold true for both males and females in view of the present findings. Furthermore, the current findings fail to support the use of intra-test
TABLE 9
ANOVA: Differences in Intra-test Scatter Associated with Sex, Test Conditions, and Anxiety Level with Runs Used as the Criterion Measure (Responses Scored 2, 1, or 0)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex(S)</td>
<td>2.043</td>
<td>1</td>
<td>2.043</td>
<td>.558</td>
</tr>
<tr>
<td>Test Conditions (T)</td>
<td>.168</td>
<td>1</td>
<td>.168</td>
<td>.046</td>
</tr>
<tr>
<td>Anxiety Level (A)</td>
<td>.043</td>
<td>1</td>
<td>.043</td>
<td>.012</td>
</tr>
<tr>
<td>S x T</td>
<td>.371</td>
<td>1</td>
<td>.371</td>
<td>.101</td>
</tr>
<tr>
<td>S x A</td>
<td>.164</td>
<td>1</td>
<td>.164</td>
<td>.045'</td>
</tr>
<tr>
<td>T x A</td>
<td>.371</td>
<td>1</td>
<td>.371</td>
<td>.101</td>
</tr>
<tr>
<td>S x T x A</td>
<td>8.172</td>
<td>1</td>
<td>8.172</td>
<td>2.233</td>
</tr>
<tr>
<td>Within</td>
<td>322.004</td>
<td>88</td>
<td>3.659</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>333.336</td>
<td>95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
scatter as a valid and useful clinical, diagnostic measure of inconsistency in intellectual test performance.

The overall ramifications of the results presented in this chapter with respect to differences in total score, internal-consistency reliability and intra-test scatter will be discussed in detail in the following chapter.
CHAPTER V

DISCUSSION

The major purpose of this investigation was to examine the effects of test anxiety upon consistency in intelligence test performance under contrasting conditions of test administration. Three hypotheses, which dealt with the effects of test anxiety (subject variables), test conditions (situational variables), and the interactions between these two independent variables, respectively, were tested. The initial hypothesis predicted that in general HA Ss (i.e., Ss who evidenced a strong predisposition to be anxious in test situations) would show greater erraticism or inconsistency in their intellectual test performance than LA Ss (i.e., Ss who did not show this predisposition), regardless of the conditions under which they would be tested. This hypothesis viewed response inconsistency as a correlate of test anxiety that would be manifested irrespective of situational variables. The first hypothesis dealt exclusively with test anxiety while in effect partialing out the effects of test conditions. In contrast, the second hypothesis dealt exclusively with the effects of differential test conditions. Implicit in this hypothesis is the notion of a reactive anxiety-like state which is induced under stressful test conditions. Accordingly, it was hypothesized that in general Ss tested under stressful conditions, irrespective of their predispositions to be anxious, would show greater inconsistency in their item-to-item responding than Ss tested under nonstressful conditions. The third hypothesis dealt with the interactive effects of predispositional
anxiety level and reaction to test conditions. Specifically, it was hypothesized that the effects of test anxiety upon consistency in performance would be most pronounced in the HA subgroup tested under stress (most inconsistency) and least pronounced (least consistency) in the LA subgroup tested under nonstressful conditions. It should also be noted at this point that the above predictions assumed that between-group differences in consistency in performance would occur regardless of whether or not differences in total score or level of performance arose.

With respect to the methodology used in testing the hypotheses, the sample consisted of 96 fifth-graders with average I.Q.'s attending public school in a predominantly white, middle-class, suburban region. A standardized questionnaire (TASC) was used to select LA and HA Ss. Ss were paired on the basis of sex, age, group I.Q. score, and test anxiety level. One member of each pair was then randomly assigned to a nonstressful treatment group; the other member to a stressful treatment group. All Ss were then individually tested with the Comprehension subtest of the WISC under either of the two alternative treatment conditions. In other words, half of the 48 LA Ss were tested under nonstressful conditions; the other half, under stressful conditions. A similar procedure was followed with the remaining 48 HA Ss.

Response Consistency in Relation to Level of Performance

Although the major interest centered upon consistency in item-to-item responding, prior to actually testing the consistency hypotheses, Ss were initially compared on total score, i.e., level of performance. This procedure was followed mainly in order to determine the extent to which consistency in item-to-item responding might be related to overall level of
performance. Using this approach, total score on the Comprehension test was used as the criterion (dependent) variable in a 2 x 2 x 2 ANOVA design. As reported earlier, there were no significant differences in total test score attributed to the effects of anxiety level, test conditions, sex or the interactions among these factors. There was also no evidence of any significant interactions involving categories or levels of factors.

Although the nonsignificant findings pertaining to total test score or level of performance do not have direct bearing upon the major hypotheses of this investigation, it should be noted in passing that these findings are not in keeping with the major trend of past research. Most recent studies, which mainly have employed correlational methods, have found significant negative correlations between scores on anxiety measures and summation scores on various intelligence tests (e.g., Alpert & Haber, 1960; Feldhusen & Klausmeier, 1962; I. G. Sarason, 1963; S. B. Sarason, et al., 1958; S. B. Sarason, et al., 1960; Walker, et al., 1970; Zweibelson, 1956). Additional studies obtaining similar findings have been cited in reviews of anxiety-related research (Kirkland, 1971; Phillips, et al., 1970; Ruebush, 1963). Hence, the majority of these investigations have suggested the possibility that anxiety is inimical to cognitive test performance in some way.

As noted earlier, however, the expected results have not always been consistently obtained. For example, in two studies dealing specifically with the WISC, scores on the Children's Manifest Anxiety Scale were correlated with scores on the WISC subscales. However, the expected results were not obtained (Hafner, et al., 1960; Rowley & Stone, 1963). It should be emphasized, however, that both of these investigations have focused upon level of performance on the intellectual measures. Neither of the above studies focused upon other aspects of test performance, e.g., consistency
in item-to-item responding, as has been done in the present investigation.

Previously it had been asserted that while differences between high and low test-anxious Ss would not necessarily be seen in level of performance, as the conflicting findings cited above might suggest, such differences would more likely be observed in the degree of consistency in their performance. This assertion is based upon the fact that predispositional anxiety, in conjunction with stress, results in inefficient and variable task performance. Evidence of this assumption was presented earlier.

The above expectation was partially confirmed in the present investigation, since the differences in the internal consistency measures are notable and suggest that these Ss did differ in consistency in item-to-item responding. However, when a nondirectional (two-tailed) test of significance was applied to the Pearson product-moment correlation coefficients for the halved tests, the difference was significant only at the .10 level. (This was not the case when a less conservative, one-tailed test of significance was used, in which case p fell below the .05 level). Thus, cautiously interpreted, the findings provide partial support for the first hypothesis. In a similar manner, some support was obtained also for the second hypothesis pertaining to situationally induced anxiety. However, the difference in the obtained Pearson r's for the halved test is significant only at the .11 level, again using a conservative, two-tailed test. The predicted effects pertaining to the interaction between the subject's anxiety level and situationally induced anxiety were also partially supported on the basis of the internal consistency reliability data.

Performance Consistency as an Aspect of Anxiety Theory

In the preceding review of the literature pertaining to anxiety theory
and research, it had been noted that anxiety may be viewed as a multi-dimensional construct (McReynolds, 1968). Along one dimension, it may be meaningful to speak of characteristic, predispositional anxiety versus situational, current anxiety (Phillips, et al., 1970). Spielberger (1966b), following Cattell and others, dichotomized this dimension into trait anxiety versus state anxiety.

The present findings conform with and, in turn, support theories similar to the state-trait theory of anxiety. This conclusion is evidenced by the fact that the predicted effects of predispositional test anxiety had occurred when the effects of test conditions were varied (Hypothesis One). It is also evident from the fact that the predicted effects of test conditions occurred when subject differences in anxiety level were manipulated (Hypothesis Two). Consistent with the results of other recent investigations (e.g., Martin & Meyers, 1972; Meyers & Martin, 1972), the findings support the notion that the effects of predispositional anxiety may be distinguished from the effects of situational, i.e., reactive anxiety. At the same time, with respect to predictions based upon Hypothesis Three, the present results do suggest that the relationship is interactive and complementary. Hence, a major implication of the present study is its support of theories which have treated predispositional anxiety and situational anxiety as separate but interactive components. But in contrast to other investigations that have focused upon level of performance, the present study has supported these theories on the basis of the consistency measures.

The view that variability in cognitive test performance is an important component of anxiety may be found in the work of Spence and Spence (1966), Mandler & Watson (1966), and S. B. Sarason (1960). A pertinent aspect of the theories of Spence and Spence (1966), which have been derived from
Hull's drive-based learning theories, is the concept of response interference and task-irrelevant behavior. Consistent with other writers, they hold that HA Ss, i.e., those who score high on an anxiety questionnaire, are apt to respond emotionally and hence inefficiently in response to complex cognitive tasks. Especially under stressful conditions, such individuals typically manifest a great number of task-irrelevant responses "reflecting self-depreciation, anger, desire to escape, etc. (p. 308)." Such behavior would be inimical to successful responding and would tend to result in variable performance. In contrast, LA Ss would be expected to show greater efficiency in their cognitive functioning and consistently achieve a higher degree of success. For such individuals mild stress may even be facilitative. Mandler and Watson (1966), discuss anxiety within the context of what they term "interruption theory." But what is pertinent to the present investigation is the fact that they also view response interference and task-irrelevant behavior as closely associated with anxiety. Sarason, et al. (1960), based upon their research dealing specifically with elementary school children, also state that anxiety is characterized by task-irrelevant behavior. To put the matter in concrete terms, these authors state that the HA child who is placed in an evaluative situation experiences many conflicting thoughts and feelings. These conflicts, according to their point of view, interfere with efficiency in test performance. While Sarason (1966) later emphasized the role of cognitive defenses against anxiety as they affect test performance, the important point is that such views do conform with the finding that the performance of the test-anxious individual is characterized by a great deal of variability.

Although there are many important theoretical differences in the ways these authors conceptualize anxiety and its behavioral correlates, Spence
and Spence (1966), Mandler and Watson (1966), Sarason, et al. (1960) all view fluctuation in attention and erraticism in performance as being one consequence (or at least a concomitant) of anxiety in test and test-like situations. In this respect, these authors, who have focused primarily upon anxiety as a research topic, would appear to be in basic agreement with previously cited clinical investigators who have presented similar views about the effects of anxiety upon cognitive performance (e.g., Rabin, 1965; Rapaport, et al., 1945; Wechsler, 1958).

Thus far, it has been suggested that in studying the effects of anxiety it may be equally or perhaps more important to focus upon the degree of erraticism in performance as opposed to S's level of performance. In addition, the findings suggest that there is an interaction between characteristic, trait-like, predispositional anxiety and situational, state-like, existent anxiety. However, there remains an additional theoretical problem to which the present findings may be related - the question of cause and effect.

The Cause - Effect Relationship Between Anxiety and Intelligence

The question of cause and effect, as it pertains to the relationship between anxiety and intelligence, arises from those correlational studies referred to earlier in which significant negative correlations have been obtained between anxiety measures and intelligence test scores. This question was dealt with by Phillips, et al., (1970). On the one hand Klausmeier and Goodwin (1966, pp. 390-393) have hypothesized that children of lower I.Q.'s have limited ability to deal with threatening situations such as test and problem-solving situations. Stated simply, in their view, children of low ability are anxious mainly because they are easily overwhelmed and often
fail at assigned tasks. In contrast, Sarason and his colleagues (1960) feel that while children do, of course, vary in ability, many high anxious children may only appear less intelligent on tests because their performance is impaired by anxiety. Furthermore, we cannot really know how intelligent such children are because their potential intelligence cannot be accurately measured, even under favorable conditions.

While the present findings do not necessarily contradict the Klausmeier and Goodwin (1966) position, they do provide greater support for the position of Sarason and his associates. In spite of the fact that there were no significant differences in level of performance that would be attributed to the effects of anxiety, the differences in the internal-consistency measures suggest that the Comprehension test did not measure the cognitive abilities of all Ss with the same degree of reliability. Accordingly, it is possible that anxious Ss may have failed many items which they might have otherwise passed had it not been for the disabling effects of anxiety. Hence, in response to the questions posed earlier, the level of performance of the HA Ss, especially those tested under stress, conceivably might have been higher but for the disruptive effects attributed to anxiety.

Although this interpretation is generally favorable to the position of Sarason's group, it does not rule out the possibility that failure on any of the items might have added to an S's predispositional anxiety level, thereby further increasing the probability of failure. This view, then, stresses the dynamic interaction that exists between anxiety and intelligence and takes into account both the positions of Sarason, et al. (1960) and Klausmeier and Goodwin (1966). Spence and Spence also interpret the relationship in dynamic terms. Their comments are especially relevant to the present investigation which dealt with both the predispositional and situational aspects of anxiety:
Instructions that stress the importance of doing well or state that performance reflects a valued characteristic, such as intelligence for the college-student subject, may be expected to lead most individuals to increased effort and attention and hence to better performance. However, emphasis on doing well may also arouse anxiety (fear of failure) and negative evaluations of performance (failure reports) to intensify it. As anxiety increases in intensity, so do the frequency and intensity of task-irrelevant responses. To the extent that the response to be acquired can be adversely affected by them, these irrelevant tendencies will lead to performance decrement. Thus, as externally manipulated psychological stress increases, performance might first be expected to increase in task-oriented behavior and then to decrease as irrelevant responses are aroused and begin to be predominant in their influence.

(With respect to predispositional anxiety), ... we might describe high anxiety Ss as having a lower anxiety threshold for the arousal of anxiety than the low anxious, tending to react even to mild ego-involving instructions with fear of failure. Thus, while the performance of low anxiety groups would be expected to rise and then decline as stress increases, the initial rise in high anxiety groups, if it appears at all, would be expected to be attenuated and their decline in performance appear not only earlier on the stress continuum but be, at any given point, more pronounced (pp. 313-314).

In addition to clarifying the relationship between anxiety and intelligence, their comments are also helpful in understanding the facilitative as well as disruptive effects of anxiety.

Anxiety, Inconsistency and the Problem of Measuring Intelligence

An important point which has been raised and which should be elaborated upon is the problem of accurately measuring the intelligence of individuals whose test performance is apt to be adversely affected by anxiety. This problem holds true regardless of whether we are referring to existent anxiety or proneness to anxiety. Along these lines, if one accepts the thesis that anxiety disrupts test performance, e.g., by negatively affecting consistency in item-to-item responding, one might ask, "Can we ever reliably measure what is sometimes conceived of as the individual's 'optimal' level of performance?" Of course, this question raises many more fundamental and complex
issues about what is really measured by intelligence tests. These issues cannot be fully dealt with within the scope of this paper. However, parenthetically, this topic does raise some methodological questions pertaining to those studies dealing with anxiety in various I.Q. groups (Feldhusen & Klausmeier, 1962; Ruebush, 1960). In this context one might question the adequacy of grouping on the basis of I.Q. level, if this measure were not accurate to begin with. Thus, it is appropriate to deal at this point with some of the overall ramifications of the present research for educational and psychological measurement.

As noted earlier, an increased interest in individual differences has highlighted shortcomings of psychological tests and psychometric theory. This topic has been the subject of numerous articles referred to earlier (viz., Anastasi, 1967; Bock & Wood, 1971; Cronbach, 1967; Kirkland, 1971; Loevinger, 1967; Mayo, 1965). In this context, the problem of intra-individual response variability had been viewed as quite significant and was the subject of a comprehensive review article by Fiske and Rice (1955). These authors, among others, have indicated the need to identify personality correlates to intra-individual response variability.

It is felt that an additional contribution of the present investigation is that it suggests a personality correlate for intra-individual response variability in its finding that inconsistency or erraticism in cognitive test performance appears to be closely associated with predispositional and/or situational anxiety level. In this sense, the findings provide support for those authors who have hypothesized that variability in test responding is associated with certain personality variables (Eysenck, 1947; Fiske & Rice, 1955; Loevinger, 1967; Sechrest & Jackson, 1967). These writers have questioned assumptions from classical test theory which have held that measurement error is a random but predictable phenomenon and that all examinees are
equally unreliable in their test performance. The present findings provide additional bases for such questioning in view of the data which suggest that certain (i.e., high anxious) individuals tend to be more unreliable in their cognitive test performance than others. However, the present findings need to be interpreted in the light of earlier findings which have direct bearing on the present investigation.

It had been stated earlier that there has been little systematic research in this area. However, Glaser (1949, 1951, 1952) is one of a few individuals to have dealt specifically and extensively with the problem. This investigator also sought to obtain empirical evidence for the notion that variability is associated with certain definitive personality characteristics and that individuals differ in their degree of consistency in performance on the basis of these personality traits. However, the findings which he obtained ran counter to this hypothesis. In contrast to the results of the present investigation, Glaser found that variability in test performance is a function of the relative difficulty of the items and the subject's overall level of performance. Furthermore, he found no evidence for the view that inconsistency is a variable associated with personality functioning.

There are several major differences between Glaser's methods and those used in the present study. As has been generally true of most previous investigations, Glaser investigated inconsistency over periods of time. His basic measure of inconsistency was based upon the number of responses changed upon retesting from pass-to-fail or from fail-to-pass. In contrast, the present study defined inconsistency as the tendency of an individual to vary in his item-to-item responding on the basis of a single-test administration. Yet, Glaser's findings do point up the need for additional research to de-
termine whether inconsistent individuals remain inconsistent over time or upon retesting. This would be particularly important in view of evidence presented by Guilford (1954) to the effect that most inconsistency measures have failed to show a high degree of reliability.

Second, although Glaser did attempt to establish a personality correlate for inconsistency, he did not deal specifically with test anxiety. This is also true of other previous investigations. In the present investigation, the theoretical and empirical bases have been presented for the assumption that individuals with high levels of predispositional or reactive anxiety will manifest a great deal of task-irrelevant behavior which is inimical to consistency in item-to-item responding. Hence, in contrast to previous investigations, the present study was specifically designed to treat as independent variables S's predispositional and induced anxiety levels.

Third, the purely cognitive tasks (which included tests of word-knowledge and computational skill) selected by Glaser may have been much less sensitive to the influence of noncognitive factors than the present tasks. It will be recalled, that in the present study a cognitive task was specifically selected for which there was some a priori evidence to assume that it would be sensitive to affective disturbance (Glaser & Zimmerman, 1967). In this connection, the extent to which item content might be pertinent is an open question in view of evidence to suggest that personality factors may also influence such seemingly neutral tasks as maze tracing (Eysenck, 1947) and various measures of scholastic aptitude (Fiske & Rice, 1955; Sechrest & Jackson, 1967). In addition to other points that have been raised, there is a need for additional research to determine also the extent to which the nature of the task may account for differences in consistency between HA and LA Ss, as well as with respect to some of
the other points that have been raised.

Diagnosis of Inconsistency

In addition to establishing that test anxiety may serve as a theoretical construct for explaining intra-individual response variability, the present investigation was motivated by an additional and more practical objective. This objective was aimed at arriving at an intra-individual measure of response inconsistency. Such a measure might be used for diagnostic purposes to differentiate highly variable individuals (in this study, i.e., those high in characteristic or current anxiety) from those who perform as would be expected, i.e., in a consistent fashion.

As was noted earlier, clinical researchers have devised various measures of intra-test scatter, or the tendency of the individual to fail easy items while passing hard items on a given point of scale. A number of studies have been reported (Hallenbeck, et al., 1965; Holzberg & Deane, 1950; Nickols, 1963; Rapaport, et al., 1945; Saxe, 1966; Watson, 1965; Wechsler, 1958). However, only limited success has been achieved in arriving at a useful diagnostic measure. In this connection, the need for more systematic research was pointed out to place such clinico-intuitive diagnostic measures on a firmer footing (Rabin, 1965; Guertin, et al., 1966).

The present investigation attempted to establish the efficacy of two measures of intra-test scatter. Both assumed that differences in the group-based variability measures (i.e., internal-consistency reliability coefficients) would be manifested at the intra-individual level by quantifying an S's tendency to give inferior responses to easier items while passing some of the more difficult items. In this connection it was assumed that variability in item-to-item responding would be evident from the number
of fluctuations in consecutive responses receiving the same score (i.e., runs) in S's performance. While the first runs measure compared Ss on dichotomously scored items, the second runs measure also took into account partial successes. However, the present investigation failed to obtain empirical support for these, clinically derived measures of intra-individual response variability. In this respect, the results of the present investigation are similar to previously unsuccessful attempts to arrive at a meaningful measure of intra-test scatter.

The repeated failure to establish a firm methodological basis for the use of intra-test scatter as a diagnostic measure reflects many of the methodological difficulties discussed in an early article (Lorr & Meister, 1941) dealing with the validity of scatter patterns derived from the Binet. These cautions would be well taken by clinicians who sometimes fail to take sufficient account of such factors as test and item specificity as well as the gradation in difficulty of the items. While many of their conclusions apply to measures of scatter derived from age scales, some of these conclusions are also applicable to point scales such as the Wechsler subscales. Hence, the present investigation would be added to the body of literature which has cautioned against the use of unsubstantiated clinical-diagnostic tests. There is a definite need for additional research in this area. Further attempts to devise valid measures of intra-test scatter might attempt to apply the present hypotheses to lengthier, homogeneous tests consisting of items that are more finely gradated in difficulty level than the items comprising the Wechsler subscales.

Conclusions

Before listing the major conclusions of this research, it is necessary to reiterate and to elaborate upon some of the factors limiting the scope
of the present investigation and to suggest areas for additional research.

One limitation pointed out earlier is the population to which the current findings may be generalized. This point is reiterated here, mainly because it is felt that one difficulty in attempting to integrate findings on anxiety as related to cognitive test performance is the heterogeneity of the populations studied. For example, there is certainly much question as to whether findings obtained on college students are applicable to elementary students, or vice versa. Likewise, research based upon middle-class, white children cannot be directly related to disadvantaged, ghetto residents. What may be anxiety provoking for dull, non-motivated students may not be for bright, achievement-oriented students. Yet even though this conclusion may appear obvious, such comparisons are often made, nonetheless. Accordingly, subject differences may account for many of the conflicting conclusions that have been presented with regard to the intelligence test performance of anxious versus nonanxious students.

It should also be reemphasized that the present findings pertain only to the effects of anxiety upon intelligence test performance as measured by the Comprehension test. As may be recalled, the Comprehension test was specifically selected because there is evidence that the content and nature of the items presumably render the test especially sensitive to affective disturbance as it might influence cognitive performance (Glasser & Zimmerman, 1967). However, as stated earlier, there is a need for additional research to determine whether anxiety would likewise influence scholastic and nonscholastic aptitude tests composed of items sampling other abilities. For example, one might reasonably ask if similar results would have been obtained had a test of vocabulary or spatial relations been used instead of a test of verbal comprehension consisting of socially relevant
items, which call for on-the-spot reasoning.

Similar questions may be directed at the TASC, as it may differ from other measures of anxiety. As mentioned earlier, anxiety assessment devices need to be carefully examined in terms of the dimensional components that are actually being measured (McReynolds, 1968). Additional insights might be provided by more extensive factor analytic studies intercorrelating existing measures of anxiety, e.g., as has been done by Alpert and Haber (1960). Such approaches would be helpful with respect to reducing some of the existing confusions pointed out earlier regarding anxiety theory and research.

Further pursuant to the problems of measuring anxiety, as it may be variously defined, is the matter of distortion and faking. Biased responding is clearly a limitation of all self-report devices. The problem has been broadly dealt with in the psychological literature.

In the present investigation, this limitation may have restricted the ability to differentiate between HA and LA Ss, in spite of the TASC's generally adequate validity. However, O'Neill and Wightman (1971) have suggested a remedy. Their application of a lie scale may enable sharper differentiation between HA and LA Ss. In addition to improving the predictibility of the anxiety scale, the use of a lie scale may provide additional data with regard to important within-group differences, i.e., those who readily admit to anxiety as contrasted with those who deny or minimize their anxieties. This topic also highlights the need for research dealing not only with anxiety but also with the psychological defenses which individuals employ to keep the effects of anxiety from conscious awareness (Ruebush, 1963; Sarason, 1966).

Another topic which requires some mention is the problem of sex differ-
ences. Within the scope of the present investigation the findings do support Kirkland's (1971) conclusion with regard to the absence of a relationship between sex and anxiety as it affects cognitive test performance in elementary school children. However, in view of conflicting findings presented by Ruebush (1963), Phillips, et al. (1970), Klausmeier & Goodwin (1966), there is need for additional research to resolve many of the issues that have bearing on this topic. In this connection, the finding that boys tend to deny their anxieties more than do girls is also worthy of further investigation (Sarason, et al., 1960).

Finally, it is necessary to comment upon the results obtained using the present statistical tests as they may limit the conclusions drawn on the basis of this research. As will be recalled, the predictions with respect to the internal consistency measures were tested under two forms of the null hypotheses. Under $H_0: \rho = 0$, the plausibility of all three experimental hypotheses was established in that comparison groups differed greatly on the obtained Pearson $r$'s. However, under $H_0: r_1 - r_2 = 0$, which may be viewed as a more direct test of the hypotheses, the experimental hypotheses would not be accepted at conventional levels of significance (viz., $p < .05$). Yet, it should be emphasized that in no case did $p$ exceed the .11 level, even with the use of nondirectional (two-tailed) statistical tests. Thus, if one allows as an acceptable risk roughly a ten percent chance of a Type I error (i.e., a probability of one-in-ten of rejecting a true null hypothesis) then the statistical results may be viewed as partially supportive of the experimental hypotheses. Hence, on this basis, several tentative conclusions may be drawn from this research. Although the findings are stated in the form of conclusions, perhaps it may be more appropriate to view these general statements as bases for extending the hypotheses for future
research. Accordingly, the findings may be summarized as follows:

1. There is some indication that a high level of test anxiety (which has been defined as the predisposition to be anxious in test and test-like situations, as expressed on the basis of a standardized questionnaire) tends to interfere with consistency in item-to-item responding on an intellectual test.

2. Stressful testing likewise appears to interfere with consistency in intellectual test performance, regardless of the individual's predisposition to be anxious.

3. In relation to test anxiety, it is probably meaningful to speak of predispositional anxiety on the one hand; and, induced, reactive, or situational anxiety on the other. This two-part conceptualization of anxiety is in basic agreement with Spielberger's trait-state theory of anxiety and with similar theories that have been proposed.

4. Although the effects of predispositional anxiety upon response consistency may be measured separately from those attributed to reactive anxiety, the findings suggest that these effects interact.

5. While the disruptive effects of test anxiety (predispositional or reactive) upon consistency in cognitive test performance may not necessarily be observed in S's level of performance, i.e., summation score, it is possible that such effects would more likely be observed on the basis of measures of consistency in item-to-item responding.

6. Internal consistency reliability coefficients may be used to compare groups of subjects that would be expected to differ in consistency in item-to-item responding, especially on the basis of differences in salient personality characteristics or on the basis of differential treatment.

7. It is suggested that the hypotheses pertaining to anxiety may pro-
vide useful constructs to account for the phenomenon referred to as intra-individual response variability.

8. The indication that some individuals may be expected to be more variable in their cognitive test performance more than others raises some research questions relative to classical assumptions concerning the reliability of intelligence tests.

9. While intra-individual variability may be regarded as a behavioral correlate of anxiety, there is a need for additional research to develop methods of differentiating inconsistent from consistent individuals on the basis of item response patterns. Such methods might eventually prove useful for making differential diagnoses between anxious and nonanxious individuals, achievers, etc.
REFERENCES


Horrocks, J. W. Assessment of behavior. Columbus, Ohio: Chas, E. Merrill, 1964.

Hoyt, C. Test reliability obtained by analysis of variance. Psychometrika, 1941, 6, 153-160.


APPENDIX A

FORMS

Contents

<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Directions for administration of TASC</td>
<td>94</td>
</tr>
<tr>
<td>2. TASC items</td>
<td>98</td>
</tr>
<tr>
<td>3. Data Collection Form</td>
<td>101</td>
</tr>
</tbody>
</table>
EXAMINER DIRECTIONS

Please familiarize yourself with the instructions and the items contained in this questionnaire so that administration may be carried out smoothly.

Previous research indicates that the questionnaire is more effective if given while the teacher is out of the classroom. Permission for this purpose should first be obtained from the school administrator and teacher. Every attempt should be made to secure the full cooperation and trust of the pupils. At the same time an orderly atmosphere should be established to ensure proper collection of the data. If possible, note any unusual circumstances or questions which might arise during administration of the TASC. Only questions which pertain to the administration and format of the questionnaire may be answered. No items should be interpreted for the pupils. Rather, students should be encouraged to indicate their initial responses as they honestly feel the questions apply to them. While individual responses may be changed, this practice should not be encouraged. Copying should be discouraged by reminding students that you are interested only in their own unique responses and that they will not be penalized in any way for their answers. Should any questions arise as to the purpose of the questionnaire, it may be appropriate to indicate only that you wish to find out how children of their age level honestly feel about school and school work. Further discussion should be deferred until collection of the papers.

The instructions should be given clearly, in a manner such that all students can hear and understand them. Standardized procedures require close adherence to instructions. However, occasional paraphrasing is permissible to allow for natural and relaxed administration of the TASC and for mainten-
ance of rapport. However, TASC items should be read aloud, *exactly as written*, while the students read them silently. It is essential to proceed at a pace which allows *all pupils to complete all items*, one by one, even if they are in doubt as to an appropriate response. This is especially important for those pupils who have difficulty or are unable to read the questions. Questions may be re-read, if necessary.
TEST INSTRUCTIONS

"My name is (Mr.) (Miss) (Mrs.) ________ I'm going to be asking you some questions - questions different from the usual school questions, for these are about how you feel. So they have no right or wrong answers. First I'll hand out the answer sheets and then I'll tell you more about the questions. Please leave them face down. Don't answer the questions until we're all ready to begin. ...

(After handing out questionnaires, say:)

"Please write your name at the top of the (first) page, both your first and your last names. Circle B if you're a boy or a G if you're a girl. Then write the name of your school and your teacher's name in the right spaces. Don't worry about the spelling. Then put down today's date in the space provided. Today is ________.

"As I said before, I am going to ask you some questions. Neither your teacher, your principal nor your parents will see your answers to the questions. These questions are different from other questions that you are asked in school. These questions are different because there are no right or wrong answers. You are to listen to each question and then put a circle around either 'yes' or 'no'. These questions are about how you think and feel and, therefore, they have no right or wrong answers. People think and feel differently. The person sitting next to you might put a circle around 'yes' and you may put a circle around 'no'. For example, if I asked you this question: 'Do you like to play ball?', some of you would put a circle around 'yes' and some of you would put it around 'no'. Your answer depends on how you think and feel. These questions are about how you think and feel about
school, and about a lot of other things. Remember, listen carefully to each question and answer it 'yes' or 'no' by deciding how you think and feel. If you don't understand a question, ask me about it. Be sure to answer all the questions. Do you all understand?

"Now let's start by everybody putting their finger on Number 1. Here is the first question. Number 1. 'Do you worry when _____?' (Repeat this procedure of introducing the questions for several of them and continue throughout to say the number of the question before reading it.)

(After question #18 is completed the examiner reads the following paragraph and continues with questions 19 - 30.)

"In the following questions the word 'test' is used. What I mean by 'test' is any time the teacher asks you to do something to find out how much you know or how much you have learned. It could be by your writing on paper, or by your speaking aloud, or by your writing on the blackboard. Do you understand what I mean by 'test' -- it is any time the teacher asks you to do something to find out how much you know."
Yes No 1. Do you worry when the teacher says that she is going to ask you questions to find out how much you know?

Yes No 2. Do you worry about being promoted, that is passing from the ___ to the ___ grade at the end of the year?

Yes No 3. When the teacher asks you to get up in front of the class and read aloud, are you afraid that you are going to make some bad mistakes?

Yes No 4. When the teacher says that she is going to call upon some boys and girls in the class to do arithmetic problems, do you hope that she will call upon someone else and not on you?

Yes No 5. Do you sometimes dream at night that you are in school and cannot answer the teacher's questions?

Yes No 6. When the teacher says she is going to find out how much you have learned, does your heart begin to beat faster?

Yes No 7. When the teacher is teaching you about arithmetic, do you feel that other children in the class understand her better than you?

Yes No 8. When you are in bed at night, do you sometimes worry about how you are going to do in class the next day?

Yes No 9. When the teacher asks you to write on the blackboard in front of the class, does the hand you write with sometimes shake a little?

Yes No 10. When the teacher is teaching you about reading, do you feel that other children in class understand her better than you?

Yes No 11. Do you think you worry more about school than other children?
Yes No 12. When you are at home and you are thinking about your arithmetic lesson for the next day, do you become afraid that you will get the answers wrong when the teacher calls upon you?

Yes No 13. If you are sick and miss school, do you worry that you will do more poorly in your school work than other children when you return to school?

Yes No 14. Do you sometimes dream at night that other boys and girls in your class can do things you cannot do?

Yes No 15. When you are home and you are thinking about your reading lesson for the next day, do you worry that you will do poorly on the lesson?

Yes No 16. When the teacher says that she is going to find out how much you have learned, do you get a funny feeling in your stomach?

Yes No 17. If you did very poorly when the teacher called on you, would you probably feel like crying even though you would try not to cry?

Yes No 18. Do you sometimes dream at night that the teacher is angry because you do not know your lessons?

Yes No 19. Are you afraid of school tests?

Yes No 20. Do you worry a lot before you take a test?

Yes No 21. Do you worry a lot while you are taking a test?

Yes No 22. After you have taken a test do you worry about how well you did on the test?

Yes No 23. Do you sometimes dream at night that you did poorly on a test you had in school that day?

Yes No 24. When you are taking a test, does the hand you write with shake a little?

Yes No 25. When the teacher says that she is going to give the class a test,
do you become afraid that you will do poorly?
Yes No 26. When you are taking a hard test, do you forget some things you knew very well before you started taking the test?
Yes No 27. Do you wish a lot of times that you didn't worry so much about tests?
Yes No 28. When the teacher says that she is going to give the class a test, do you get a nervous or funny feeling?
Yes No 29. While you are taking a test do you usually think you are doing poorly?
Yes No 30. While you are on your way to school, do you sometimes worry that the teacher may give the class a test?
1. Cut - Finger

2. Lose - Ball

3. Loaf - Bread

4. Fight

5. Train - Track

6. House - Brick

7. Criminals

8. Women - Children

9. Bills - Check

10. Charity - Beggar

11. Gov't - Examinations

12. Cotton - Fiber

13. Senators

14. Promise - Kept

Raw Score

Scaled Score

Runs
The dissertation submitted by Lawrence H. Zaiden has been read and approved by members of the Department of Foundations, School of Education.

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

5/18/72
Date

Signature of Advisor