THE INFLUENCE OF METHOD OF ADMINISTRATION, SEX OF EXAMINER, AND SEX OF SUBJECT ON THE RELIABILITY AND VALIDITY OF THE MARIANNE FROSTIG DEVELOPMENTAL TEST OF VISUAL PERCEPTION

MILDRED FRANCES ST. CROIX

THE INFLUENCE OF METHOD OF ADMINISTRATION, SEX OF EXAMINER, AND SEX OF SUBJECT ON THE RELIABILITY AND VALIDITY OF THE MARIANNE FROSTIG DEVELOPMENTAL TEST OF VISUAL PERCEPTION

by

© Mildred Frances St. Croix

A Thesis presented to

The Department of Educational Psychology

Memorial University of Newfoundland

In partial fulfillment

of the requirements for the degree

Master of Education

April 1976

St. John's Newfoundland

ABSTRACT

The purpose of this study was to investigate the reliability of the Frostig Developmental Test of Visual Perception (FDTVP), the effects of the sex of examiner, the sex of examinee, and mode of administration. The FDTVP was administered by twenty examiners to 123 subjects stratified according to demography (rural or urban), sex (male or female), grade (I or II), and mode of administration (group or individual) in a test-retest design. Examiner sex/subject sex combinations were changed from one testing occasion to the other for each subject. Based on the results of a study of the inter-rater reliabilities, protocols were randomly assigned to four persons for scoring.

Percentile rank distributions of the total sample verified that the mean scores obtained in this study were very high relative to the norms established by Frostig.

The overall test-retest reliability of the FDTVP was lower than reported by Frostig. The test-retest reliability findings for the group administration were generally higher than those found for individual administrations. Grade II reliabilities were much lower than Grade I reliabilities. For Grade I, the group administration produced the higher reliabilities.

The analysis of variance showed few differences for sex of subject, sex of examiner, and administration mode.

In all cases, the raw scores for each subscale, and the sum of scaled scores were analyzed. The sex of subject was significant only for subscale I. For subscale IV, a significant two-way interaction was found for sex of subject and sex of examiner. For the same subscale a significant three way interaction was found for sex of subject, sex of examiner and administration mode.

Findings suggest that Grade II reliabilities are too low to warrant any use of the test for diagnostic or remedial prescriptions. For Grade I children, the group setting is the indicated administration mode. In addition, all existing evidence on the validity of the FDTVP must be questioned because of the nature of the score distributions.

ACKNOWLEDGEMENTS

This study is the result of the wise counsel and the generous contributions of a number of persons.

To Dr. William Spain, my supervisor, my sincere gratitude for his advice and assistance. His criticisms, suggestions and confidence have guided my efforts in this study from their formative stages to the concluding statements.

My thanks to Dr. Lorne Taylor for helping to clarify problems of the on-going study.

To the principals and the schools who participated in this study, my sincere appreciation for their help and support.

I owe a debt of gratitude to the many generous persons who assisted in collecting my data.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

STATEMENT OF PURPOSE

The purpose of this study was to compare the reliability and validity of the Frostig Developmental Test of Visual Perception (FDTVP) under individual and group modes of administrations. In addition, the influence of the sex of the examiner, and the sex of subject was studied.

SIGNIFICANCE OF STUDY

"Reliability is a necessary but not sufficient condition for validity" (Stanley and Hopkins, p. 114). It is important to ask about the precision or reliability of a measure because one must generalize from the observation in hand to some class of observations to which it belongs. Cronbach, Rajaratnam, and Gleser (1963) characterize a good measuring instrument as having three qualities: validity, reliability, and usability. This statement echoes throughout the literature as a basic tenet in statistical research (Linquist, 1953; Cronbach, 1962; and Thorndike, 1968).

Kerlinger (1965) states: ". . . concern for reliability comes from the necessity for dependability in measurement" (p. 429). It is generally accepted that the data of all psychological and educational measurement instruments contain

errors of measurement. To the extent that they do, the data they yield will not be valid and therefore any conclusions drawn from them will not, of course, be valid.

One important theme underlying the theoretical question of reliability involves the identification of possible sources of unreliability in the measurement instrument. Administrative procedures, time limits, and instructions are among the primary sources of possible error (Cronbach, 1962; Anastasi, 1965; and Stanley and Hopkins, 1972). Entering into this category is the problem of whether the test is designed for individuals or group administration.

Standardization of individual and group tests implies different procedures in administering and interpreting the tests. If the scores obtained by different methods of administrations are to be comparable, the test constructors must provide detailed directions for the administration of the test (Anastasi, 1965). This standardization must, in theory, extend to the exact materials employed, time limits, oral instructions to subjects, preliminary demonstrations, ways of handling queries about items, and other details of the testing situation. Another important step in the standardization of a test for the group or the individual is the establishment of different norms, thus providing for a difference in the interpretation of results (Anastasi, 1965).

The advantages and disadvantages of group and individual tests have been discussed in theory. The claim is

made that group tests provide the opportunity and convenience of mass testing, simplification of the examiner's role, objective scoring, and better established norms (Anastasi, 1965). The advocates of individual testing claim that individual tests provide an opportunity to establish rapport, obtain cooperation, maintain the interest of subjects, and to identify temporary conditions of the subjects such as illness, fatigue, or anxiety, which could have a bearing on the test results (Linquist, 1953; Guilford, 1956; Anastasi, 1965; and Keough and Smith, 1967). In short, the dyadic relationship of the individual testing situation should result in less measurement error because of the ability of the examiners to intervene directly in testing.

Most standardized tests are designed either to be administered to groups or to individuals (Linquist, 1953; Anastasi, 1965). However, a few test constructors claim that their tests may be administered interchangeably in either way. The Bender-Gestalt Test, the Embedded Figures Test, and the Marianne Frostig Developmental Test of Visual Perception are cited as examples of double administration tests. There is no evidence to date of any research on the contrasting reliabilities of those specific tests under group and individual administrations. There are, however, some validity studies reported which show consistent differences between group and individual administration modes (Keough and Smith, 1961; Jackson, 1964). The study by Keough and Smith on the

Bender-Gestalt Test indicated no significant differences between the modes of administration. Correlations between group and individually administered Embedded Figures Test indicated sufficiently high agreement to warrant substitution of group for individual forms (Jackson, 1964).

It is apparent from the preceding discussion that from theory, there is no way of predicting the comparability (validity and reliability) of a test under group or individual administrations.

The constructors of the FDTVP claim that the test is equally valid and reliable under either group or individual administration. However, there is no evidence presented in the test manuals or in the literature reviewed which would support such a claim.

A study quoted by Frostig, Lefever, and Whittlesey (1960) using an individual administration, same examiner, and a three-week interval between test-retest, reports a product-moment correlation of .98. A second study (Frostig et al., 1961) using group administration, two examiners, and a two-week interval, reports a correlation coefficient of .80. It can be inferred from this data that assuming comparability of samples, an assumption which does not seem warranted, approximately 20 per cent of the score variance of the FDTVP is due to the examiner and to the mode of administration; however, this variance cannot be partitioned as to source. The evidence strongly supports the hypotheses that mode of

administration, or the examiner, or both are significant sources of error variance in administering the FDTVP.

The literature also suggests that sex of examiner and sex of subject are related to the level of performance on tasks similar to those presented by the FDTVP. Consistent findings of sex differences in testing are reported as mean differences (Melker and Garfield, 1951; Masling, 1960; and Harris, 1971), and as variability differences (Stevenson, 1961; Stevenson, Keen, and Knights, 1963). The Stevenson et al. (1963) study reports females with more variable responsiveness in scoring than males. These findings suggest the ability to control motivation of subject in individual and group testing may be partly a function of sex of examiner. Given that the difference between group and individual administration methods is the nature of the relationship between examiner and subject, it is possible that the sex of the examiner and subject will interact with administration mode to show differences in reliability and validity.

It would appear that since the FDTVP is used fairly extensively to make both screening and placement decisions, the question of effect of mode of administration on these decisions becomes important. If, as is suggested by the data, 20 per cent of the score variance of the FDTVP is due to error of measurement, serious placement decisions can be made by using reliability findings which do not consider the effects of administration mode.

This study has attempted to study the reliability and validity of the FDTVP using stratification of the variables therefore partitioning the sources of error variance.

HYPOTHESES

1. There are no differences in the test-retest reliabilities obtained from group administrations and individual administrations of the FDTVP.
2. There are no significant differences in mean scores obtained from group administrations and individual administrations of the FDTVP.
3. There are no significant differences in mean scores obtained from male and female subjects on the FDTVP.
4. There are no significant differences in mean scores obtained from male examiners and female examiners on the FDTVP.
5. There are no significant differences in mean scores obtained from Grade I and Grade II subjects on the FDTVP.
6. There are no significant differences in mean scores obtained from the first testing occasion and the retesting occasion of the FDTVP.
7. There is no significant interaction of sex of subject and sex of examiner on the FDTVP.
8. There is no significant interaction of sex of subject and administration mode on the FDTVP.

9. There is no significant interaction of sex of subject, sex of examiner, and administration mode on the FDTVP.
10. There is no significant interaction of sex of examiner and administration mode on the FDTVP.

LIMITATIONS OF THE STUDY

There exists in this study, as there must in any study of this nature, limiting factors the reader must bear in mind when appraising the results and conclusions. The following limitations are noted:

(a) The generalizability of the findings will be limited to populations similar to the sample studied; and

(b) the findings based on the Frostig Developmental Test of Visual Perception may not be applicable to all visual perceptual tests but only to those with tasks similar to the FDTVP.

CHAPTER II

REVIEW OF THE LITERATURE

I. INTRODUCTION

This chapter is a review of the literature relevant to the study, and presents the conclusions which evolve from that review. The reviews are divided into three segments which are not independent of each other but rather are complementary to the overall purpose of this study.

Section II presents studies on the examiner-subject interaction. Section III discusses the factors influencing the measurement of visual perception. Section IV selects one visual perceptual test, the Frostig Developmental Test of Visual Perception, and reviews the research on that test.

II. EXAMINER EFFECTS

Research in experimental psychology has shown somewhat striking results which indicate that examiners (Es) may and do influence their data (Kintz et al., 1965). McGuigan (1963) states: "While we have traditionally recognized that the characteristic of an examiner may indeed influence behavior, it is important to observe that we have not seriously attempted to study him as an independent variable"

(p. 42). Stumpf, an influential scientist of Germany, began the study of the examiner as an independent variable in 1904, but not until recently has the problem been considered by experimental psychologists for study (Cordaro and Ison, 1963; McGuigan, 1963). Kintz (1965) contends that wherever an experimenter-subject relationship exists, the possibility also exists for E to contaminate his data by one or more of a multitude of conveyances.

Two contentions recur in the literature: (i) that an E biases a subject's (S's) responses either directly as a result of his personal characteristics, for instance, one's sex or personality, or indirectly through the nature of the relationship; and (ii) the E's interpretation of Ss' responses may be biased, for instance, in scoring (McGuigan, 1963; Kintz et al., 1965).

This section presents research findings on examiner influences under the categories of personality, sex, expectancies, interpretations, for instance, in scoring, and within the context of his relationship with a subject.

Personality Variables

In attempting to assess effects of personality factors of Es in the experimental situation, McGuigan (1960) compared trait scores of Es on personality tests with dependent variable scores of Ss. However, no significant results were obtained. Sanders and Cleveland (1953) using

projective technique tests found that overtly anxious Es tended to elicit more subject flexibility and responsiveness than covertly anxious Es on the Rorshach instrument.

Rosenthal, Persinger and Fode (1962) reported that not only Es' personality but the Ss' perception of this personality can contribute to the E effect. Investigation of Ss' perception of E has been undertaken by two related studies (Rosenthal et al., 1960; and Rosenthal and Persinger, 1962). Both studies supported the hypothesis that naive Ss may have a predetermined "set" about what a typical E is like, for instance, scientific or intelligent. Lord (1950) and Masling (1957) maintain that an E, by acting "warm" or "cold" can influence a subject's responses to a projective test.

These studies presented seem to indicate that the personal characteristics of the examiner play a role in influencing either the behavior of the examiner and/or the behavior of the subject.

Sex Influences

Several studies have been concerned with investigating the manner in which results are influenced by examiner differences in sex. Carlson and Carlson (1960) commented that sex of subject and sex of examiner is an omnipresent but largely ignored variable. The sex of E and S influences such diverse phenomena as conformity and acquiescence

(Crutchfield, 1955; Beloff, 1958; Tuddenham, 1958), need affiliation (Exline, 1962), empathy (Dymond, 1950; Tomlinson, 1967), classroom learning (Page, 1958); verbal conditioning (Binder, McConnell and Sjoholm, 1957; Buss and Durkee, 1958; Sarason and Minard, 1963), and E bias (Gewirtz, 1954; Stevenson, 1961; Rosenthal, 1966; Haan and Livson, 1973). Recent studies (Pheterson, 1969; Horner, 1970; Silverman et al., 1972; and Piacente, 1974), investigating the evaluations of Es as a function of their sex and competency, show that results vary with sex of examiner.

Masling (1960) warns that sex of E and S are among the numerous situational factors influencing psychological test performance. Harris (1971) likewise points out the necessity of noting the sex of examiner and subject when doing psychological research.

Curtis and Wolf (1951) and Clark (1952) reported that sex of E influenced the production of responses on projective type tests. Gewirtz and Baer (1958) and Stevenson (1961) reiterate those conclusions. Stevenson (1961) offers explanations of why this is so. He says the Oedipal theory of Freud provides a meaningful focus for his results, for instance, female adults may have a significantly greater effect on the performance of boys than of girls in the age range of four to seven years, and male adults have a significantly greater effect on the performance of girls than of boys for the same age range (Stevenson, 1961). In

another study, Stevenson, Keen, and Knights (1963) found that the least effective combination was of male examiner with a male subject, and the most effective combination was a female examiner with a female subject. Results obtained in this study were explained in terms of anxiety levels raised by male examiners compared to those raised by female examiners.

Stabler (1967) explained his findings in terms of differing social interactions involved between Es and Ss. He found that male Es succeeded in having their Ss respond to their authoritative commands whereas females were viewed as having a minor role in authority. Similarly, Gray (1959) reported that identification with an examiner of the same sex is psychologically more important for boys than for girls. Gray, for example, found indices of male child identification with a male E were significantly correlated with the adjustment of boys, but that the relationship did not hold for girls with female Es. Stevenson (1961) hypothesizes that the effectiveness of social reinforcement provided by an adult examiner is a function of the degree to which children are deprived of contact with members of the adult's sex.

Masling (1969) and Harris (1971) are among many researchers (Brunswick, 1956; McGuigan, 1963; and Klein, Cicchetti, and Spohn, 1967) who are concerned over research conducted ignoring the possible influence of sex of E and Ss on results. Harris (1971) states that as a minimum step to

control for the effects of sex, one can report the number and sex of Es and Ss, thus allowing others to replicate the study. Also, if sex of E has a differential effect on Ss, opposite sex Es can expect differences in their data when they attempt to replicate each other's work (Harris, 1971). When these sex differences appear as interactions (Masling and Harris, 1969), rather than simply as main effects (Klein et al., 1967), awareness of sex becomes particularly important. Brunswick (1956) says that ignoring the possible influence of sex allows a source of error variance into one's experiment that could be eliminated. McGuigan (1963) states that broad sampling of the examiner population permits the psychologist to place more confidence in his conclusion that differences are due to one's sex and not to the idiosyncrasies of one man and one woman.

Jackson (1964) obtained results showing data for male subjects revealed more differentiated performances in measures of perceptual speed and spatial orientation than did data for female subjects. This finding of a different pattern of correlations among cognitive performances in the two sexes is consistent with previous research (Jackson, 1955; Bieri et al., 1958; Witkin et al., 1962; Messick and Kogan, 1964) and suggest once again the fruitfulness of further explorations of the differential conditions out of which cognitive differentiation emerges in males and females.

Three studies which have reported evidence of examiner variability associated with sex and race of examiner in the administration of the Stanford-Binet are of relevance. Cieutat (1965) reports a significant difference between the scores elicited by female and male Es on the Stanford-Binet (Forms L-M). The female Es obtained significantly higher scores than male Es when testing minority group children ranging in age from 47 to 52 months. The overall mean of the six female examiners was 89.61, and that of the male examiners was 83.19.

Cieutat and Flick (1966) present additional evidence of E variability on the Stanford-Binet (Forms L-M) scores of 448 four year old Negro children. Marginally significant interaction was reported between the sex of the examiner and sex of subject. Analysis of individual test items indicated that examiner sex bias was present in most cases where the test items were of medium difficulty.

Forrester and Klaus (1964) report a study of 25 five and six year old Negro kindergarten children. Sex of examiner was found to be a significant variable on the 1937 revised Stanford-Binet.

Le Bovitz (1966) verified the findings of Cieutat and Flick (1966). Le Bovitz suggested that the primary source of bias is likely to be related to examiner characteristics and includes factors such as sex.

Expectancy Effect

Perhaps the component of examiner effect which is the cause of greatest concern is that by which the E in some way influences the Ss to perform as the E has hypothesized (Kintz et al., 1965). The reasons for concern about expectancy effect are because so little is known about it and so little research appears to be devoted to it.

Several studies were found, however, regarding the expectancy effect. Rosenthal and Fode (1963) demonstrated the problem clearly in an experiment with two groups of randomly assigned animals. This study concluded that examiner expectancies of subjects were influencing factors on subjects' performance, for example, the expected 'bright' subjects performed significantly better than the expected 'dull' subjects. Allen and Feldman (1974) support results obtained by Rosenthal and Fode (1968) that examiners' expectations of subjects will affect subjects' performances.

Simon (1969) exposed a different effect of examiner expectancies. He clearly demonstrated that expectations will affect the scoring process. He maintains that the motivational properties of the expectancies formulated by any given individual in any given situation are a function not only of the E's own particular personality organization, but are also quite obviously a function of the relationship between the E and subject(s).

Bauer (1975) clarifies Simon's (1969) suggestions and offers possible explanations why expectancies have direct influences on responses and interpretations of those responses. Bauer maintains that expectancies based upon impersonal stimulus inputs may be reinforced by information obtained from personal interaction with teachers, parents, or with the subject himself. Secondly, an examiner might possess a somewhat different expectancy for a subject who has done exceptionally well on the first part of a test than for a subject who has done rather poorly on the same fraction of the text (Bauer, 1975).

The Mode of Administration

The preceding survey of the literature on examiner influences would not be complete if the nature of the working relationship is ignored, for instance, whether the study is carried in a group or an individual setting.

The literature review to date evidences the importance of the relationship between examiner and subject. However, few studies are devoted to the prime purpose of studying effects under differing relationships.

Masling (1959) addresses the problem under the area of intelligence testing. He says:

> During the course of their training most testers are exhorted to establish "rapport" and admonished to be "objective." The "objective" examiner is charged with the responsibility of deriving as valid an estimate of the intelligence of the subject as

can be obtained, without regard for his personal attitudes about the subject. He is thus expected to be standardized and depersonalized.

Masling's study attempted to investigate the extent to which an examiner could divest himself of personal bias in administering and scoring an intelligence test when the examiner acted in either a highly interested ('warm') manner in an individual testing session or in a persistently rejecting, disinterested ('cold') manner. Masling's (1959) results indicated the subjects' responses were better where the examiner had a direct control over the relationship in the dyadic set-up.

Gordon and Durea (1948) found that a group, given the Stanford-Binet under conditions of a large group, informal setting, earned a mean IQ score 6.35 points lower than a smaller group.

Hutt (1960) concluded that poorly adjusted children score better on performance tests in a setting where examiner-subject interaction is high as opposed to informal group situations.

Bauer (1970), in line with Hutt's study (1947), noted that a possible source of error in scores obtained from differences in test instructions demonstrated that test-anxious boys were dependent on how the examiner structured his relationship with the boys. Test-anxious boys apparently rely on sources of control outside themselves for directions, support, and approval.

All studies (Hutt, 1947; Gordon and Durea, 1948; Masling, 1959; and Bauer, 1960) reiterate the conclusion that subjects' performances suffer, not because of lack of potential on the subjects' part but because of the level of interaction between examiner and subject.

Several studies of a different nature, regarding the group and/or individual working relationship, appear in the literature. Those studies reflect on the specific details of an individual experiment which differ from those involved in a group experiment. These studies will be cited and the conclusions will be presented in an attempt to clarify this other aspect of the working relationship.

Jackson (1964) discusses the advantages and disadvantages of individual and group tests. His contentions are in line with those expressed by Guilford (1956), Linquist (1953) and Anastasi (1965) that research on individual differences requires large samples if any confidence is to be placed in the replicability of results in generalizing to a particular population. Also, it is often extremely difficult and costly to enlist hundreds of subjects for the period of time required for extensive individual testing. However, past attempts to construct group administered tests requiring different instructions from those used in individual tests have met with some success according to reported studies (e.g. Thurstone, 1944; Guilford and Lacey, 1947; and Overlade, 1956).

The relevance of the effects of a differing working relationship on the measurement of visual perceptual tasks is important in this study. Studies were found in this area and will be quoted mainly because of their findings and implications for this present study.

Jackson (1964) conducted his investigations on the Embedded-Figures Test. A specific recommendation of this study pointed towards the necessity of differing instructions for a group mode of administration than for an individual mode. For a group administration, for instance, each subject must be in direct alignment with the stimulus object so that the maximum different orientation of the design is possible and such that the presentation is identified for all subjects.

Fuller (1969) and Lasch et al. (1974) in studying the Minnesota Percepto-Diagnostic Test add to the detailed instructions involved in differing the mode of administration, for instance, light reflecting off the screen may impair results, especially those at acute angles to the screen and difficulty in perceiving the figures. Lasch et al. (1974) postulated that stimulus cards could be kept in direct alignment with the copying paper during a group administration in order to convert the Minnesota Percepto-Diagnostic Test to a group test from an individual test. Howard (1970) further comments on the previously discussed studies with her work on the Bender-Gestalt test. She demonstrated the

processes which occur when subjects must alternate eye focus between near and distant stimuli, for instance, from the screen and the paper.

The studies (Jackson, 1964; Howard, 1970; and Lasch et al., 1974) are quoted as examples of studies which have noted the differences of group and individual administrations. All studies do recognize the different instructions required and maintain while some differences may involve reliable and consistent effects they might prove neither necessary nor relevant to the primary aim in test construction, yet others may prove indispensable.

Jackson (1964) stresses the implications of either a group setting or an individual setting saying that it is possible that the dyadic interaction may have effects that transcend the primary purpose of the test, for instance, a subject's performance may be altered due to his embarrassment over failing an item.

Scorer Differences

The entire field of scorer differences has received attention by researchers of psychological testing. Evidence of scorer reliability differences on the Bender-Gestalt Test have been reported (Miller et al., 1963; Werner, 1966; Egeland, Rice and Penny, 1967; and Broadhurst and Phillips, 1969). In general, the reported interscorer reliabilities vary between .79 and .96, and on the whole exceed .85; thus

indicating the test's interscorer reliability is "adequate."

However, Egeland, Rice, and Penny (1967) reported one of their scorers on the Bender-Gestalt Test was more lenient than the other two. They hypothesized that such a variation among trained scorers would lead one to expect even more variation among 'naive' or inexperienced scorers of the test. Similarly, Broadhurst and Phillips (1969) tried to explain why one of their scorers deviated markedly from the other three. Thus, the question was raised whether the degree of clinical experience and handling of the test influences interscorer reliability.

Morsbach et al. (1971) attempted to answer this question. The findings suggested that interscorer reliability coefficients were generally sufficiently high in both the "experienced" and "inexperienced" scorer groups. However, the untrained group was much more homogeneous in its scoring than the clinical group since the latter group tended to split into two subsections, one with rather severe scorers and the other with two rather lenient scorers. Morsbach (1971) found that the reliability of the test was sufficiently high in both groups concluding that questions about the reliability of the test seemed to be due less to variations of scoring inside the scorers and more to variations between scorers.

The Weschler Intelligence Scale for Children (WISC) has received treatment in the area of differential scoring

(Glasser and Zimmerman, 1967; Madden, 1974; and Rothman, 1974). Madden (1974) states: "Directions for the scoring of the mazes subtest of the WISC, as given in the manual (Weschler, 1949) are ambiguous."

Excluding deliberate departures from established procedures, for instance, over or undertiming, variability may be caused by such factors as verbal approval (Witmer et al., 1971) and examiner expectancy of success (Dickstein and Kephart, 1972). Both studies were done on the WISC.

Three other studies (Satler and Wiget, 1970; Miller and Chansky, 1972; and Rothman, 1974) report differential vulnerability in the WISC particularly in Vocabulary, Similarities, and Comprehension.

From the literature reviews, it can be seen that examiner differences can be elicited in the form of scorer difference due to a wide number of conditions; for example the ambiguity of scoring instructions by test constructors.

Implications for Testing

The preceding survey of the literature has revealed the existence of the experimenter effect in several aspects of psychology. Postman and Jarrett (1952) comment:

> We have paid too little attention to the contributions made by variations in E's behavior to the examiners results. The difficulty which many researchers experience in repeating the results of other investigators may be due to our failure to attack systematically the role of differences among Es. (p. 253)

The general area of testing which would include IQ tests, perception tests, aptitude tests, reading readiness tests, among others, seems to suffer if the examiner subject influences are ignored. Even though there has been rigorous attempts at standardization of test items and procedures, the examiner still influences the test taken in other subtle ways (Kanfer, 1958; and Rosenthal, 1963).

It is often questionable whether many tests have been proven sufficiently reliable and valid in their own right, and this examiner variable further complicates the issue. Judgment of an individual's score on a test must not only be viewed in light of which test was used, but must also take into consideration the previously ignored variable of the specific examiner and the working relationship. All persons using test scores must recognize the strong influence of the examiner and make decisions accordingly (McConnel, Bender and Sjoholm, 1957; McGuigan, 1963; and Rosenthal, 1963).

III. MEASUREMENT OF THE VISUAL PERCEPTUAL FACTORS

Age--The Developmental Factor

Fidel and Ray (1972) state: "Between the ages of three and seven years, perceptual development becomes an extremely sensitive indicator of the general developmental status of the child." Synder and Synder (1974) in a study of maturational changes in visual perception on the Bender-

Gestalt showed rapid maturing skill from age six to seven years and that level of difficulty decreased considerably with chronological growth.

Goulet (1974) in a series of concentrated studies investigating the effects of age, school experience and the development of visual perception, used a battery of tests, for example, the Frostig DTVP, the Visual Memory Test, the Visual Motor Gestalt Test, the Illinois Test of Psycho-linguistic Abilities, the Visual Sequential Memory Test, and the visual subtest from the Reading Aptitude Test. He obtained results suggesting that the interaction effects indicated a greater change in six months for the kindergarten children in comparison to any other grade.

Lasch et al. (1974) supported the results obtained by Goulet (1974) and further claims that experiments based upon visual perception ought to have sampling defined strictly according to age rather than grade since "ages may vary within a given class due to retention, double promotion, etc." Potter (1966) showed a correlation between age and efficiency of perceptual recognition. Henning and Kornreich (1971) attempted to replicate and extend Potter's study (1966). The latter study obtained results showing a marked developmental effect on the recognition test for the children of ages three to seven years.

Practice Effects

Potter (1966) and Henning and Kornreich (1971) also studied practice effects as related to perceptual recognition tasks. Both studies concluded that practice had an effect only on the youngest children, that is, the nursery and kindergarten ages. For those children, the results of Neuman Keuls' test revealed that the tracing group did significantly better than the other groups. Thus, the tracing effect was pronounced in the young children but vanished with older children.

This conclusion was earlier reported in a similar experiment by Zaporozhets (1965) and supports the educational theory and practices of Montessori educators (1964).

Studies on General Perceptual Factors

The effects of method of measurement upon test perception performance of children with various exceptionalities have been repeatedly demonstrated. Goldstein (1948) and Sarason (1953) have shown that brain-injured children are significantly inferior in their ability to translate percepts into words when compared to nonbrain-injured children. Cruise (1961) found that brain-injured children are more distractible than nonbrain-injured children. Bortner and Birch (1960) found that although brain-injured children are able to perceive geometric forms, they are unable to reproduce them accurately.

Newcomer and Hammill (1973) demonstrated the effects of a motoric method of measurement upon performance by administering the Motor Free Test of Visual Perception and the Bender-Gestalt Test to children with varying degrees of motor handicap. The fine motor coordination requirements in the method of measurement for the Bender-Gestalt test depressed performance and yielded an unreliable estimate of visual-perceptual skills (Ritter and Sabatino, 1974).

Werner and Strauss (1941) found that brain-injured children were significantly inferior to nonbrain-injured children in performance on visual figure-ground perception tasks. In a follow-up study, Rubin (1969) heeded the finds of distractibility as a possible intervening factor and with slight modification of the method of measurement procedures found no difference in the performance of the two groups.

A study (Ritter and Sabatino, 1974) suggests that there may be a greater contribution of method variance to scores than is often assumed. The primary implication for the practitioner is that tests are measuring more than they purport to measure, and that caution should be exercised in the inference of skill deficiency because of poor perceptual performance.

Ritter and Sabatino (1974) support previous hypotheses (Corah and Powell, 1963; Becker and Sabatino, 1973) of visual figure-ground perception and form discrimination as

being essentially the same.

Domrath (1968) assessed the role of visual perception in the performance of certain constructional praxis tasks by second grade children. A "perceptual" hypothesis followed the views expressed by Mayer Gross and Denny-Brown that constructional activity is essentially a motor expression of the visuo-perceptive field. However, Kleist (1953) raised an alternate hypothesis that constructional apraxia results from an impairment of the linkage between the visual percept and the corresponding motor performance.

IV. THE FDTVP

Introduction

> To facilitate the early detection and categorization of such perceptual impairment, a preliminary instrument was devised and normative data obtained so that the degrees and kinds of deviations from these norms could be assessed. On the basis of the results of testing a sample of 434 normal children and a sample of 71 children, ages 3½ to 8 years, a new version of this test was developed. (Frostig, Lefever and Whittlesey, 1961, p. 383)

The above quote serves as part of the rationale put forth by the authors of the Marianne Frostig Visual Test of Perception (FDTVP). The FDTVP is composed of five subtests: Test I: Eye-Motor Coordination; Test II: Figure-Ground Relationships; Test III: Constancy of Shape; Test IV: Position in Space; and Test V: Spatial Relations.

In that article (Frostig et al., 1961) the authors emphasized that FDTVP's areas of perception are in accordance

with the findings of Cruickshank et al. (1957) and Hammill, Colarusso and Wiederholt (1970) trace the evolution of the FDTVP and claim that the authors of FDTVP integrated the works of Thurstone (1944), Cruickshank (1957) and Wedell (1960) to arrive at the finished product.

The authors claim that the FDTVP is suitable for group or individual administration. An individual administration can be completed in 30 to 45 minutes and a group from 45 to 60 minutes. Frostig (1963) states: "Scoring is objective, and requires 5 to 10 minutes" (Frostig Manual, p. 467). In her article (Frostig et al., 1961) she says: "When scoring was standardized, the inter-judge reliability of trained judges was found to be high, .90 or above" (p. 386).

Research on FDTVP

Frostig et al. (1961) claim that the subtests do measure distinct areas of visual perception. Corah and Powell (1963), Sprague (1963), Ohnmacht and Rosen (1967), Allen (1968), Crawley et al. (1968), Ohnmacht and Olson (1968), Olson et al. (1968), Boyd and Randle (1970), and Hammill et al. (1970) have conducted factor analytic studies on the independence of the FDTVP subtests. These researchers employed a variety of intelligence, readiness, achievement, and other perceptual tests, in addition to the FDTVP subtests. Regardless of the particular measures included in

these factor analyses and of the types of subject samples, all studies have failed to identify five separate perceptual factors (Hammill, 1970). However, two exceptions found the subtests of the FDTVP loaded on two perceptual factors, the two being Corah and Powell (1963) and Crawley et al. (1968).

Ward (1970) examines both the Corah (1963) and the Crawley (1968) study exposing some limitations of both, e.g., the small heterogeneous samples used. Two other studies, Silverstein (1965) and Ward (1970) added to the controversy about the validity of subtest scores as predictors of reading achievement. Silverstein (1965) and Ward (1970) concluded from their factor analytic studies that a number of distinct areas of perception can be delineated in the Frostig, FDTVP, and Achievement Studies.

The literature reports studies done on the Frostig perceptual quotient and measures of reading achievement (Bryan, 1964; Maslow, Frostig, Lefever and Whittlesey, 1964; Cohen, 1966; Olson, 1966; and Tauber, 1966).

This investigator found four studies that dealt with the predictive validity of the Frostig subtests. Olson (1966) found that Frostig Subtests I, II, IV and V were correlated significantly with second-grade achievement for third-graders, Subtest III was the best predictor of reading achievement. Cohen (1966) reported that Subtest III may have special significance in relation to reading achievement of the socially disadvantaged children in his

sample, since it correlated most highly with reading of all the perceptual and intelligence tests that he administered. Similarly, Wiederholt (1973) claimed that Frostig's confidence in Spatial Relations as a predictor of academic skills was justified.

Mlodnosky (1972), in a study using the Bender-Gestalt and the Frostig tests as predictors of first grade reading achievement among economically deprived children, found that economically deprived children score considerably lower than the Bender and Frostig standardization samples. Mlodnosky (1972) found that Subtests I and II are not good predictors of reading success. Mlodnosky supports Olson's (1966) premise that those two subtests are predictive of later reading or that these are related to skills, such as handwriting, that were not measured by the studies conducted. Mlodnosky (1972) questions Subtest III and suggests that Subtest III measures an ability to pay attention and follow complicated instructions in addition to a visual perceptual skill. Subtest V, according to her study, needs further research in order to be a better indicator of reading success. Thus, Mlodnosky's study concluded with serious reservations about the Frostig test as a good predictor of reading success. Leiburt and Shark (1970), in a study using average subjects in Grades I, II, and K, support Mlodnosky (1972) that performance on each subtest and reading success are uncorrelated on the FDTVP.

Smith and Marx (1972) suggest that the FDTVP measures a single general factor of perceptual organization which is weakly related to IQ and unrelated to reading ability; this study used the FDTVP, the WISC, and a reading test.

Wiederholt (1973) investigated the predictive validity of the FDTVP using kindergarten and first grade economically deprived subjects. His results indicated the Frostig's confidence in Spatial Relations as a predictor of academic skills was justified. Frostig (1961) postulated that of her five subtests, Spatial Relations would be the most significant predictor of writing and reading ability; Wiederholt (1973) agrees. Wiederholt's results did not support Frostig's hypothesis regarding eye-hand coordination, figure ground and form constancy, however. Wiederholt recommends that the FDTVP should be revised due to the inability of some of the subtests and of the total raw score to demonstrate useful levels of prediction.

Braithwaite (1972) and Goulet (1974) conclude that the magnitude and significance of the correlation between the FDTVP and reading achievement test scores decrease as grade level increase.

Braithwaite (1972) summarizes his findings on the relationship between the Frostig Training Program (devised by the authors of the FDTVP as special training program based on the test construction and used for subsequent

improvement in academic performance, Frostig (1961, p. 394)), and reading achievement. His conclusions are that children whose perceptual quotient is low, that is, under 90, will gain in visual perceptual skills after completion of the Frostig Perceptual Training Program if, and only if, the FDTVP is the criterion measure. He further cautions that improved visual perception skills through completion of the training program will not solely ensure success in reading (Braithwaite, 1972).

Reliability Studies

The only reliability studies reported on the Frostig are those reported by Frostig et al. (1963).

The first study (1960) was conducted using an individual administration, same examiner, and a three-week interval between test-retest reports a product-moment correlation of .98.

The second study by Frostig (1961) was done using a group administration, two examiners, and a two-week interval reports a correlation coefficient of .80.

It can be inferred that approximately 20 per cent of the score variance is due to examiner and mode of administration since a correlation of .80 was obtained with no partitioning of the source of error accounted by examiners and/or administration mode.

V. SUMMARY

A survey of the literature pertinent to examiner variables, perceptual factors, and the Frostig TVP has been presented. While the research is relevant to the problem of this investigation, several points emerge:

1. Examiner, subject, and examiner subject interactions--especially sex--have been demonstrated to influence testing outcomes.
2. Differences between scorers and scorer expectancies influence the variability of scoring.
3. The testing situation, especially group versus individual administration, has shown to be differential for some tests.
4. Examiner and subject will interact with situation.
5. No studies of these factors were found for the FDTVP. Reliability studies reported suggested that some of these factors were important.

CHAPTER III

GENERAL PROCEDURE

A sample of 123 primary school age students was selected, stratified by grade, geographic location of the school, and sex. The subjects within each level of stratification were randomly assigned to either the individual or group mode of administration. The FDTVP was administered to each subject on two occasions, separated by a minimum of seven and a maximum of ten days. The examiners were assigned randomly on each occasion, however ensuring that the sex of the examiner was stratified so that each subject had a male and female examiner, and that there were equal numbers of subjects examined by male examiners and female examiners at each level of stratification on each occasion of testing.

SAMPLE

Table III.1 shows the number of subjects in each level of stratification. Frostig has three-month-age groups in her norming procedures. Table III.2 presents the breakdown of those age groups for this study. It can be observed from Table III.2 that the sample used in this study were normally distributed according to the age ranges suggested by Frostig. Chi-square findings on randomization are also included in Table III.2.

TABLE III.1

FREQUENCY DISTRIBUTION OF PUPIL SAMPLE GROUPED ACCORDING TO ADMINISTRATION MODE, GRADE, GEOGRAPHIC LOCATION, AND SEX

Admin. Mode	Urban Center				Rural Center				Total
	Grade I		Grade II		Grade I		Grade II		
	M	F	M	F	M	F	M	F	
Group	7	8	8	8	8	8	8	8	63
Individual	7	7	8	8	6	8	7	9	60
Sex Totals	14	15	16	16	14	16	15	17	
Grade Totals	29		32		30		32		
Geographic Totals	61				62				
								N	123

TABLE III.2

AGE RANGE BY ADMINISTRATION MODE AND GRADE

Age	Grade	Administration		Total
		Group	Individual	
6-0 to 6-2	I	7	7	14
6-3 to 6-5	I	6	9	15
6-6 to 6-8	I	12	8	20
6-9 to 6-11	I	4	6	10
7-0 to 7-2	II	7	5	12
7-3 to 7-5	II	8	10	18
7-6 to 7-8	II	8	9	17
7-9 to 7-11	II	8	9	17
Total		63	60	123

Chi-square = 3.98 with 8 degrees of freedom; $p > .05$.

SPECIFIC PROCEDURES

A test-retest procedure was followed. The twenty examiners, of whom ten were male and ten female, administered the FDTVP to the 123 subjects on two occasions. Caution was taken to ensure that all possible sex combinations of examiner and subject were equally represented. On retesting, the examiner sex/subject sex combinations were changed for each subject.

Administrative Details

The group administration time length was a maximum of one hour. The individual administration took from 30 to 45 minutes.

All administrations occurred between 9:30 A.M. and 2:30 P.M. All testing took place during the month of December, 1975.

The time interval between the first administration and the second administration was a minimum of seven and a maximum of ten days. According to Frostig et al. (1963) test-retest correlation coefficients tend to be low if retesting is done after a long interval, that is, more than two weeks, or if testing is done by trained non-psychologists ($r = .69$), or if training is intervened between testings. Frostig (1963) contends that this is partly due to the rapid development of visual perception, which appears to be more highly correlated with age than with IQ or with achievement.

SCORING

An inter-rater reliability study was conducted prior to the total scoring of the test. This was a pilot study using forty randomly chosen test protocols, each scored by four different examiners. The identity of the protocol was withheld from the scorers. All scores were examiners used in this study.

The results of the inter-rater reliability study are summarized in Table III.3. Differences existed in the reliabilities obtained on the five subtests. The reported reliabilities of the perceptual quotient (PQ) as .78 and the percentile rank (PR) as .69 seem to be low.

TABLE III.3

INTER-RATER RELIABILITIES OF FDTVP SCORES

Sub Scales	Raw Score	Age Equivalents
I	.83	.81
II	.74	.70
III	.97	.98
IV	.94	.90
V	.82	.82
Sum Scaled Scores	PQ	PR
.93	.78	.69

As a direct result of those findings, the investigator decided to have several people score the tests with two stipulations: (a) each scorer must have administered the test; and (b) a scorer of the first testing occasion could not score the retest of that same individual.

There were seven different scorers for the total sample. Tests were randomly assigned to each scorer. Scorers were instructed to follow the directions outlined in the Marianne Frostig manual (1963, pp. 17-29).

STATISTICAL ANALYSIS

Reliabilities were assessed in terms of the correlations of the test-retest scores for the total sample, grades, administration mode, and for combinations of the latter two variables. In testing hypotheses about reliabilities, direct comparison of the reliability coefficients was the procedure followed. A statistical test of the differences between these reliability coefficients was not available. Kristoff (1974) points out that little work has been done to determine the sampling distributions of test-retest reliability coefficients.

Validity findings were reported in terms of analysis of variance for the main effects and the interaction effects of sex of subject, examiners, occasion of testing, and location of testing.

All null hypotheses were rejected at the .05 level of confidence. Percentile rank distributions were found for the total sample and for each grade to supplement the analyses.

CHAPTER IV

ANALYSIS OF THE DATA

Hypothesis 1: There are no differences in the test-retest reliabilities obtained from a group administration and an individual administration of the FDTVP.

Table IV.1 presents the data on the test-retest reliabilities for this hypothesis.

From the table, it can be seen that with the exception of subscale III, the group administration scores resulted in higher reliability coefficients than the individual administration scores. Subscale I reported the largest differences with a correlation for group testing of .41 and the individual testing correlation of .28.

TABLE IV.1

TEST-RETEST RELIABILITIES FOR GROUP AND INDIVIDUAL

Admin. Mode	No.	RAW I	RAW II	RAW III	RAW IV	RAW V	SUB-SCALE	AEQ I	AEQ II	AEQ III	AEQ IV	AEQ V	PQ	PR
Group	61	.41	.78	.71	.39	.46	.72	.42	.76	.63	.33	.32	.54	.52
Ind.	62	.28	.39	.74	.20	.41	.72	.29	.37	.72	.21	.40	.62	.63

Hypothesis 2: There are no significant differences in mean scores obtained from group administrations and individual administrations of the FDTVP.

Table IV.2 shows a comparison of mean scores with analysis of variance data obtained from both group and individual administrations for subscale raw scores and the sum scaled score.

TABLE IV.2

ANALYSIS OF VARIANCE: ADMINISTRATION MODE

	Group		Individual		F-
Mode	Mean	Variance	Mean	Variance	Value
I	17.12	13.745	18.11	9.311	3.8876
II	18.67	3.616	18.30	4.191	.4027
III	11.58	11.908	11.025	10.709	1.0008
IV	7.21	.937	6.835	1.526	.6782
V	6.54	.920	6.42	1.115	.6782
SSS	55.33	36.401	55.39	39.698	.8064

There were no significant differences on any of the variables tested at the .05 level of confidence. The null hypothesis was not rejected.

Hypothesis 3: There are no significant differences in mean scores obtained from male subjects and female subjects on the FDTVP.

Table IV.3 presents the analysis of variance data for male and female subjects for subscale raw scores and the sum of scaled scores.

TABLE IV.3

ANALYSIS OF VARIANCE: SEX OF SUBJECT

Sub-Scales	Males		Females		F-Value
	Mean	Variance	Mean	Variance	
I	17.11	14.005	18.13	9.015	4.1665*
II	18.60	3.584	18.38	4.267	.5025
III	11.21	14.195	11.39	8.559	.1077
IV	6.99	1.485	7.05	1.047	.1343
V	6.41	.954	6.54	1.079	.7246
SSS	55.56	38.256	56.16	38.102	.3280

* $p\ (1,107) \leq .05$

As can be seen in Table IV.3, only subscale I was significant at the .05 level of confidence. The null hypothesis was rejected for subscale I; however, it was not rejected for the other subscales.

Hypothesis 4: There are no significant differences in mean scores obtained from male examiners and female examiners on the FDTVP.

Table IV.4 presents the analysis of variance for male and female examiners for subscale raw scores and the sum scaled scores.

TABLE IV.4

ANALYSIS OF VARIANCE: SEX OF EXAMINER

Sub-Scales	Male Examiner		Female Examiner		F-Value
	Mean	Variance	Mean	Variance	
I	17.71	9.886	17.53	13.633	.2468
II	18.35	3.933	18.63	3.903	2.1340
III	11.17	12.457	11.43	10.281	1.0803
IV	6.98	1.379	7.07	1.150	.6000
V	6.55	.862	6.41	1.171	1.8319
SSS	55.68	37.714	56.04	38.755	.7146

There were no significant differences on any of the variables tested at the .05 level of confidence. The null hypothesis was not rejected for any of the subscales.

Hypothesis 5: There are no significant differences in mean scores obtained from Grade I subjects and Grade II subjects on the FDTVP.

Table IV.5 shows the analysis of variance data for the grade variable for subscale raw scores and the sum scaled score.

TABLE IV.5

ANALYSIS OF VARIANCE: GRADE I AND GRADE II

Sub-scale	Grade I		Grade II		F-Value
	Mean	Variance	Mean	Variance	
I	16.49	11.566	18.54	10.045	20.6970*
II	18.09	4.386	18.80	3.378	5.745*
III	10.47	12.583	11.93	9.413	8.0186*
IV	6.77	1.803	7.23	.747	9.0736*
V	6.31	1.023	6.61	.974	3.9136*
SSS	58.43	42.133	53.40	22.573	29.337*

* $p\ (1, 107) \leq .05$.

The F values obtained in every case were above the critical limit indicating significant differences between grades. The null hypothesis was rejected for all subscales.

Hypothesis 6: There are no significant differences in mean scores obtained from first occasion of testing and the retest occasion on the FDTVP.

Table IV.6 presents the analysis of variance data on occasion of testing for sum scaled raw scores and the sum scaled score.

TABLE IV.6

ANALYSIS OF VARIANCE: OCCASION OF TESTING

Sub-scales	Test I		Retest		F-Value
	Mean	Variance	Mean	Variance	
I	17.97	11.778	17.07	11.517	8.2963*
II	18.07	4.818	18.84	2.899	21.4160*
III	10.56	11.345	11.83	10.918	36.3170*
IV	6.91	1.252	7.09	1.388	2.2210
V	6.30	1.242	6.62	.751	10.2415*
SSS	55.28	40.782	56.56	35.764	10.9107*

* $p\ (1,107) \leq .05$.

As can be observed in Table IV.6, significance was obtained at the .05 level for subscales I, II, III, V, and for the sum scaled score. The null hypothesis was rejected for subscales I, II, III, V, and for the subscale score; however, it was not rejected for subscale IV.

Hypothesis 7: There is no significant interaction of sex of subject and sex of examiner on the FDTVP.

Table IV.7 presents the analysis of variance data for interaction of sex of subject and sex of examiner.

TABLE IV.7

ANALYSIS OF VARIANCE: SEX OF SUBJECT AND SEX OF EXAMINER

Subscale	ME-MS X	ME-FS X	FE-MS X	FE-FS X	F-Value
I	17.35	16.87	18.06	18.19	.7283
II	18.33	18.86	18.37	18.39	1.7873
III	11.11	11.32	11.24	11.56	.3798
IV	6.82	7.17	7.14	6.97	4.8404*
V	6.43	6.39	6.66	6.43	.8632
SSS	55.11	56.01	56.24	56.07	1.5436

* $p\ (1,119) \leq .05$

Except for subscale IV, no significant interactions were found between sex of subject and sex of examiner. The null hypothesis was not rejected for subscales I, II, III, V, and for the sum scale score; however, it was rejected for subscale IV.

Hypothesis 8: There is no significant interaction of sex of subject and administration mode on the FDTVP.

Table IV.8 presents the analysis of variance data on the interaction of sex of subject and administration mode for the subscale raw scores and for the sum scaled score.

TABLE IV.8

ANALYSIS OF VARIANCE: INTERACTION OF SEX OF SUBJECT AND ADMINISTRATION MODE

Sub-scale	Group		Individual		F-Value
	ME	FE	ME	FE	
I	17.17	17.09	18.24	17.97	2.9636
II	18.59	18.75	18.11	18.50	.5425
III	11.55	11.61	10.80	11.25	1.5007
IV	7.10	7.33	6.86	6.81	2.2515
V	6.64	6.44	6.45	6.38	.1464
SSS	56.08	56.58	55.28	55.51	.5615

The null hypothesis was not rejected for the subscale raw scores for the sum scale score since no significance was found at the .05 level of confidence.

Hypothesis 9: There is no significant interaction of sex of subject, sex of examiner, and administration mode on the FDTVP.

Table IV.9 presents the analysis of variance data for the subscale raw scores and for the sum scale score for this hypothesis.

TABLE IV.9

ANALYSIS OF VARIANCE: SEX OF SUBJECT, SEX OF EXAMINER, AND ADMINISTRATION MODE

Sub-scale	Group				Individual				F-Value
	ME-MS	ME-FS	FE-MS	FE-FS	ME-MS	ME-FS	FE-MS	FE-FS	
I	16.77	17.57	15.71	18.47	17.93	18.56	18.04	17.91	3.6748
II	18.45	18.73	18.87	18.63	18.21	18.00	18.86	18.15	.9686
III	12.00	10.21	11.18	11.77	10.21	11.38	11.18	11.32	.3798
IV	7.10	7.10	6.89	7.20	6.54	7.18	6.89	6.74	4.8403*
V	6.55	6.73	6.35	6.53	6.32	6.59	6.43	6.32	.8632
SSS	56.26	55.90	56.42	56.73	53.96	56.59	55.60	55.41	4.0810*

* $p\ (1,119) \leq .05$.

Significance was obtained for subscale IV and for the sum scaled score. No other subscales were significant at the .05 level of confidence. The null hypothesis was rejected for subscale IV and for the sum scaled score; however, it was not rejected for the other subscales.

Hypothesis 10: There is no significant interaction of sex of examiner and administration mode on the FDTVP.

Table IV.10 shows the analysis of variance data on the interaction of sex of examiner and administration mode for the subscale raw scores and for the sum scaled score.

TABLE IV.10

ANALYSIS OF VARIANCE: INTERACTION OF SEX OF EXAMINER AND ADMINISTRATION MODE

Sub-scale	Group		Individual		F-Value
	ME	FE	ME	FE	
I	17.17	17.09	18.24	17.97	.7007
II	18.59	18.75	18.11	18.50	.3840
III	11.55	11.61	10.80	11.25	.6387
IV	7.10	7.33	6.86	6.81	.2762
V	6.64	6.44	6.45	6.38	.3358
SSS	56.08	56.58	55.28	55.51	.9342

All variables examined were not significant at the .05 level of confidence. The null hypothesis was not rejected.

SUPPLEMENTARY DATA FOR THE TOTAL SAMPLE

The means obtained in this study for the overall subscale raw scores are high for some subscales in comparison to Marianne Frostig's maximum possible score. Table IV.11 presents a comparison of Frostig's maximum scores and the mean scores obtained in this study.

TABLE IV.11

COMPARISON OF MEAN SCORES BY GRADE AND MAXIMUM POSSIBLE SCORES

	I	II	III	IV	V
Maximum Score	30	20	17	8	8
Grade I	16.49	18.10	10.5	6.8	6.3
Grade II	18.5	18.8	11.9	7.2	6.6
Total Sample	17.5	18.5	11.8	7.0	6.5

In order to investigate the possible impact of score distributions on the study, a distribution of percentile ranks was obtained for the total sample. Table IV.12 presents this data.

It can be observed from Table IV.12 that the distribution of scores is negatively skewed. The majority of the scores fall at the upper end of the scale with 26.8 per cent of the scores falling at or above the 95 percentile. The

distribution cannot be viewed as a continuous distribution because of the large number of scores at the upper end of the score distribution suggesting a truncation of the theoretical distribution for this sample. The percentile ranks obtained in this study tend to be much higher than expected based on Frostig's norms. Since the percentile ranks are based on the subscale raw scores, it can be inferred that these, too, are higher than expected.

TABLE IV.12

DISTRIBUTION OF PERCENTILE RANKS FOR TOTAL SAMPLE WITH DESCRIPTIVE STATISTICS

Percentile Rank	Relative Freq.	Cum. Freq.	Freq.
$\leq$ 28	1.6	1.6	2
29 - 50	12.0	13.6	15
51 - 75	22.1	25.7	27
76 - 93	37.5	63.2	46
$\geq$ 94	26.8	100.0	33

Mean	Var.	Mode	Mdn.	Range	Skewness
77.4	394.9	95	84.9	72.5	-1.05

SUPPLEMENTARY DATA FOR GRADES

Since the FDTVP claims to be developmental, it seemed beneficial to make a comparison of the test-retest reliabilities of younger and older children's performances. For this purpose, the reliabilities of Grade I and II subscale raw scores, age equivalent scores, perceptual quotients (PQ), and percentile ranks (PR) were reported. Table IV.13 presents this data:

TABLE IV.13

TEST-RETEST RELIABILITIES FOR GRADES

		Raw						AEQ						
Grade	No.	I	II	III	IV	V	SSS	I	II	III	IV	V	PQ	PR
I	59	.52	.69	.78	.30	.26	.75	.52	.63	.73	.29	.21	.68	.68
II	64	.09	.27	.65	.19	.57	.53	.11	.31	.58	.22	.51	.47	.50

As can be observed in the table, correlation coefficients obtained for Grade II, except for subtest V, are consistently lower than coefficients reported for Grade I between test and retest reliabilities of the FDTVP.

To further complement the data on the grade variable, a breakdown of percentile ranks was done. Table IV.14 presents this data:

TABLE IV.14

DISTRIBUTION OF PERCENTILE RANKS FOR GRADES WITH DESCRIPTIVE STATISTICS

Percentile Rank	Grade I Cum. Freq.	Grade II Cum. Freq.	I Freq.	II Freq.
≤ 30	1.7	4.7	1.7	4.7
31 – 60	13.6	23.4	13.6	23.4
61 – 78	13.6	17.1	13.6	17.1
79 – 93	35.5	36.0	35.5	36.0
≥ 94	35.6	18.8	35.6	18.8

Grade	Mean	Var.	Mode	Mdn.	Range	Skewness
I	81.57	309.9	95	91.4	67.5	-1.35
II	73.64	448.9	95	80.2	72.5	-0.80

Both distributions are negatively skewed; this finding is in agreement with that for the total sample. However, Grade I reports a distribution which exceeds Grade II in its skewness. Grade I has 35.6 per cent of its scores falling at or above the 95 percentile and Grade II has 18.8 per cent falling in this region. Grade I with a range of 67.5 reports scores less variable than Grade II with a range of 72.5. Grade II, showing the greater variability, elicited the lower reliability coefficients.

Since differences were observed between grades for reliabilities, it was considered helpful to compare reliabilities of the breakdown of grade by administration mode. Table IV.15 reports the reliability coefficients for subscale raw scores, age equivalent scores, perceptual quotients (PQ), and percentile ranks (PR).

TABLE IV.15

TEST-RETEST RELIABILITIES: GRADE BY ADMINISTRATION MODE

Grade	Admin. Mode	No.	Raw I	Raw II	Raw III	Raw IV	Raw V	Raw SSS	AEQ I	AEQ II	AEQ III	AEQ IV	AEQ V	PQ	PR
I	Group	30	.74	.80	.81	.50	.38	.78	.70	.78	.73	.44	.25	.72	.69
	Ind.	29	.09	.59	.75	.16	.15	.70	.24	.49	.72	.12	.18	.63	.65
II	Group	31	-.14	.72	.59	.14	.61	.45	-.01	.74	.49	.17	.49	.38	.39
	Ind.	33	.34	.29	.72	.21	.62	.62	.23	.24	.71	.24	.61	.59	.64

According to Table IV.15, test-retest correlation coefficients reported for Grade I showed individual administration scores consistently less reliable than group administration scores for all variables. Subscale I, for instance, reported a group coefficient of .74 and an individual coefficient of .09 for Grade I. There are marked differences in reliability related to administration mode for Grade I.

Table IV.15 shows variable reliability results for Grade II on the breakdown of administration mode. Subscale I, for instance, yielded a low negative correlation of -.14 for the group administration and a coefficient of .34 for the individual administration. However, on subscale II, the group administration coefficient of .72 was much higher than the individual coefficient of .29 for Grade II. Except for subscale II, the correlation coefficients for group administrations were much less reliable than those for individual administrations.

In interpreting the significance of the differences between reliability coefficients, the differences in group size should be taken into account. Reliability coefficients based on samples of size sixty will be more stable than those based on samples of size thirty.

CHAPTER V

DISCUSSION

Certain observations have been made on the findings of this study which reflect on their interpretations and conclusions drawn. Percentile rank distributions of the total sample (Table IV.12) verified that the mean scores obtained in this study were very high relative to the norms established by Frostig. Sixty-four per cent of the total sample scored at or above the 65th percentile, as can be observed in Table IV.12. The distribution is negatively skewed and can be considered truncated and therefore may not be continuous.

RELIABILITY

The theoretical upper limit of the test scores can be considered to be truncated. This would tend to make the total score distributions more homogeneous than would be expected. The high subscale scores would tend to be more homogeneous than expected as well. Reliability may be defined as a relationship between obtained score variance and error variance of measurement (Thorndike, 1964); that is, reliability = $1 - \frac{s^2_e}{s^2_0}$. From this, it follows that if

error variance of measurement is constant, then homogeneous scores will be less reliable than more heterogeneous score distributions.

The effect can be observed by noting that perceptual quotients and percentile ranks are less reliable than the sum of scaled scores, the score from which they are derived. Perceptual quotients and percentile ranks are established based on the distribution of sum scaled scores for Frostig's norming sample. This establishes an upper limit on the distribution of sum scaled scores since all scores above the 95th percentile, regardless of their size, are scored as 95th percentile. A similar process occurs in evaluating subscale scores. The result is that the total score variance will be more homogeneous for samples which perform better than Frostig's norming sample. Some support for this proposition can be found in a study by Zimmerman et al. (1968). They prove that theoretically, the classical formulation for reliability given above holds for the reliabilities of two different groups only when the means of the groups are the same. Because of the need to use Frostig norms, this assumption was focused upon the investigation even though it has been demonstrated not to be the case.

Hypothesis 1: <u>There are no differences in the test-retest reliabilities obtained from a group administration and an individual administration of the FDTVP.</u>

Table IV.1 reported the findings for this hypothesis. Group administration scores were more reliable than individual administration scores for all subscales with the exception of subscale III. This finding is consistent with evidence in the literature that group administrations may provide more reliable indications of learning assessments at early ages (Howard, 1970; Henning and Kornreich, 1971).

The reliability coefficients obtained in this study were not consistent with those reported by Frostig (1963). Perceptual Quotient reliabilities reported by Frostig for an individual administration was .98 and for a group administration was .80. Reliabilities for perceptual quotients in this study were much lower than Frostig's with a coefficient of .62 for individual administrations and .54 for group administrations.

An inconsistency was noted in the reliabilities reported for this hypothesis based on perceptual quotients. While the subscale raw score reliabilities for group administrations were higher than those for individual administrations, the perceptual quotient reliabilities were reversed. This inconsistency may be explained by the effect of the truncation of the distribution earlier described.

Supplementary Reliability Data

Table IV.13 reported test-retest reliabilities for grades. With the exception of subscale V, Grade I reliabilities were considerably higher than Grade II reliabilities.

This finding can be viewed in terms of the developmental nature of the test. The age range in Grade I was six years to six years eleven months, and in Grade II was seven years to seven years eleven months. There is evidence in the literature supporting a high positive correlation between visual perceptual growth and chronological growth (Potter, 1966).

According to Synder and Synder (1974), the critical age for development of visual perception is from five and one half years to seven years. This suggests that the true score variance of older students would be more homogeneous. The observed score variance of Grade II students was actually larger than for Grade I students. Therefore, the lower Grade II reliabilities would indicate that the portion of observed score variance attributable to error would be larger in Grade II than in Grade I.

Table IV.15 reported the findings on reliabilities for the breakdown of grade by administration mode. Group administration reliabilities exceeded individual administration reliabilities for all subscales in Grade I. The reverse results were obtained for Grade II showing individual administration reliabilities greater than group administration reliabilities for all subscales with the exception of subscale II. Those findings lend support to a hypothesis put forward in the literature that younger children are more reliable on a group relationship than older children (Howard,

1970; Henning and Kornreich, 1971).

This is a reasonable interpretation of the findings of this study if one hypothesizes a significant socialization effect in Grades I and II, leading students to be more comfortable with a variety of adult authority figures.

VALIDITY

Hypothesis 2: There are no significant differences in mean scores obtained from group administrations and individual administrations of the FDTVP.

Table IV.2 reported the findings for this hypothesis. The hypothesis was accepted based on the results.

All mean scores were very high which indicated that the theoretical truncation of the scales possibly suppressed differences which may have been present, making interpretation of the effects of this variable impossible.

Hypothesis 3: There are no significant differences in mean scores obtained from male subjects and female subjects on the FDTVP.

Table IV.3 showed the findings for this hypothesis. The hypothesis was accepted for all subscales with the exception of subscale I. Subscale I is a measure of eye-motor coordination. Females mature faster than males with respect to this characteristic (Stevenson, 1961), so it appears that this finding is consistent with research findings in the literature.

Hypothesis 4: There are no significant differences in mean scores obtained from male examiners and female examiners on the FDTVP.

The findings for this hypothesis may be found in Table IV.4. The null hypothesis was accepted. The failure of this study to find raw score differences related to the sex of the examiner does seem at variance with the findings in the literature (Masling, 1960; Cieutat, 1965; and Stabler, 1967). Once again, a possible explanation for this finding might be found in the insensitivity of the FDTVP to differences due to the truncation of the distribution and the homogeneity of the samples.

Hypothesis 5: There are no significant differences in mean scores obtained from Grade I subjects and Grade II subjects on the FDTVP.

This null hypothesis was rejected; findings may be found in Table IV.5. Grade II students obtained higher mean scores than Grade I students. The developmental nature of the FDTVP is the most probable explanation for this finding. It should be observed, however, that while significance was expected, the raw score differences are quite small suggesting that the developmental characteristic measured by the FDTVP did not change remarkably between grades in this sample. Two possible hypotheses arise as a result of this observation. First, it is possible that visual perceptual

motor development slows down at this time. Second, perhaps the FDTVP does not validly reflect visual perceptual motor development at these ages.

Hypothesis 6: There are no significant differences in mean scores obtained from first occasion of testing and the retest occasion on the FDTVP.

The null hypothesis was rejected for all subscales with the exception of subscale IV. The most probable explanation for this finding may be found in practice effects from the first to the second testing occasion.

Hypothesis 7: There is no significant interaction of sex of subject and sex of examiner on the FDTVP.

As reported in findings in Table IV.7, the null hypothesis was accepted for all subscales with the exception of subscale IV. Table V. 1C shows a graphic representation of the interaction effects of sex of examiner and sex of subject for subscale IV. According to this table, female Es elicited the higher responses from male subjects; whereas the lower responses were elicited from male subjects with male examiners. Subscale IV involves the discrimination of reversals and figures presented in series; basically, a non-motoric response is tested in this subscale. The question arises why not with subscales I, II, III, and V? Further research may be needed on this subscale to warrant any detailed conclusions.

Hypothesis 8: There is no significant interaction of sex of subject and administration mode on the FDTVP.

As reported in the findings, this null hypothesis was accepted.

Hypothesis 9: There is no significant interaction of sex of subject, sex of examiner, and administration mode on the FDTVP.

The null hypothesis was accepted for all subscales with the exception of subscale IV and for the sum scaled score. This same finding has been reported on Hypothesis 7. It is appropriate to note that subscale IV involves a non-motoric response. Tavle V. 1A and 1B presents a graphic picture of those interaction effects. It can be observed in Table V. 1A that male examiners elicited consistent responses from both male and female subjects in the group administrations. Female examiners elicited the highest responses from male subjects in the group administration. Those findings are consistent with evidence reported in the literature (Masling, 1960; Harris, 1970).

Table V. 1B shows the interaction effects of sex of examiner and sex of subject for the individual administration. Female examiners elicited higher responses from male subjects than from female subjects. However, the largest discrepancy occurred between male examiners for sex of

subject. Male examiners elicited the poorest responses from male subjects and the highest responses from female subjects in the individual administration.

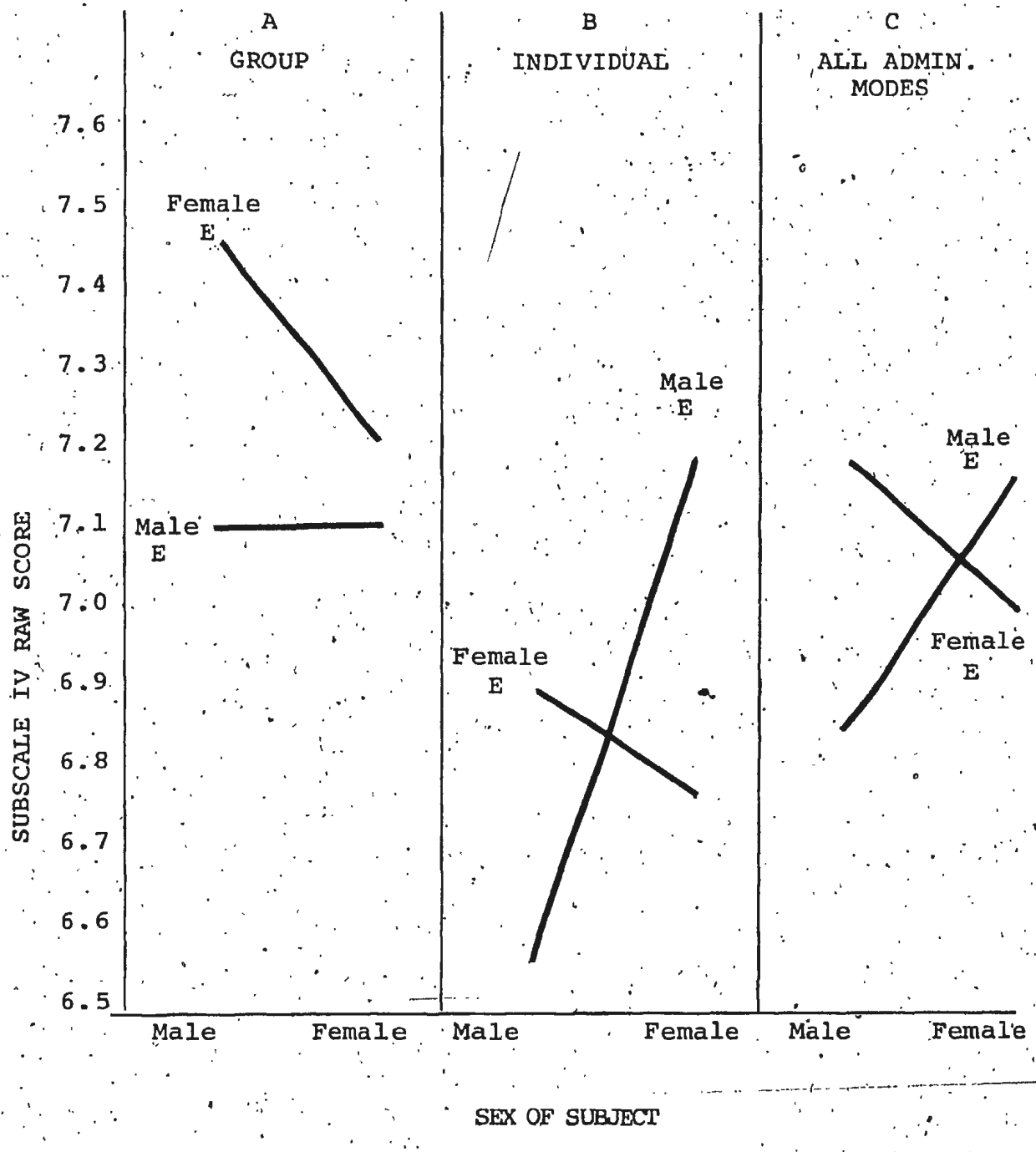
In comparing the overall interaction effects of sex of subject and sex of examiner by administration mode, female examiners elicited the highest responses from the male subjects in the group administration. Male examiners elicited the second highest responses from female subjects in the individual administration. All findings are consistent with research reported in the literature (Masling, 1960; Harris, 1970; and Golden, 1974).

Hypothesis 10: <u>There is no significant interaction of sex of examiner and administration of the FDTVP.</u>

The null hypothesis was accepted for this interaction effect.

TABLE V. 1

INTERACTION EFFECTS: SEX OF EXAMINER BY SEX OF SUBJECT BY ADMINISTRATION MODE



CHAPTER VI

SUMMARY AND IMPLICATIONS OF THE STUDY WITH RECOMMENDATIONS

SUMMARY

The purpose of this study was to compare the reliability and validity of the FDTVP under individual and group modes of administration. In addition, the influence of the sex of the examiner and the sex of subject was studied.

A sample of 123 primary age students was selected, stratified by grade, geographic location of the school, and sex. The subjects within each level were randomly assigned to either the individual or group mode of administration. The FDTVP was administered to each subject on two occasions, separated by a minimum of seven and a maximum of ten days.

The examiners were assigned randomly on each occasion, however, ensuring that the sex of the examiner was stratified. Each subject had a male and female examiner, and there were equal numbers of subjects examined by each sex of examiner at each level of stratification on each occasion.

A pilot study was conducted on the inter-rater reliabilities of the scoring of the tests. Due to the variability of scores, the investigator decided to use several scorers for the total sample.

Validity was studied using analysis of variance techniques. Reliability findings were reported in correlation coefficients.

FINDINGS

Mean scores obtained for all subscales on all variables were quite high relative to the maximum possible scores for subscales. Reliability coefficients obtained seemed low. Reliabilities for group administrations were higher than reliabilities for individual administrations for all subscales with the exception of subscale III which was only slightly lower. Correlation coefficients obtained by grades differed with Grade I reliabilities reported as higher than those for Grade II with the exception of subscale V. A breakdown of grade by administration modes showed Grade I group scores as considerably higher than Grade I individual scores; Grade II group scores, however, were lower than Grade II individual scores, with the exception of subscale II.

The analysis of variance showed very few differences. In all cases, the raw scores for each subscale, and the sum of scaled scores were analyzed. There were no significant differences due to the mode of administration and sex of examiner. The sex of subject was significant only for subscale I. Significance was obtained for all subscale scores for the difference between Grades I and II. Occasion of

testing yielded significance for all subscales with the exception of subscale IV. Analysis of variance showed significance only for subscale IV on interaction effects between sex of subject and sex of examiner; between sex of subject and administration mode; between sex of examiner and administration mode.

IMPLICATIONS OF THE STUDY

General implications of this study might be:

1. The use of the FDTVP as a screening device for normal children would lead to large numbers of errors in screening decisions because the low reliabilities would lower the ability of the test to discriminate among children with varying levels of visual perceptual motor development.

2. Reliabilities in Grade II are so low that the test is virtually useless in discriminating among children with varying levels of visual perceptual motor development.

3. Subscale scores at all levels tend to be so low that they would lead to extremely high levels of placement error if they are used diagnostically to place children in perscriptive remedial programs.

4. All existing evidence on the validity of the FDTVP must be questioned because of the nature of the score distributions.

5. For Grade I children, the group setting is the indicated administrative mode. If administrered in Grade II

the indicated mode is individual.

RECOMMENDATIONS FOR FURTHER STUDY

While the reliability and validity of the FDTVP for Grades I and II was explored in this study, it is recommended that kindergarten and Grade III samples be studied. The slightly higher raw scores obtained by Grade II over Grade I introduces the question of developmental effects which could be answered in the extremities of the age/grade range were included.

The need for a close examination of Frostig et al.'s (1963) norms is suggested by the high mean scores obtained for all subscales.

It is further recommended that a study be conducted to determine the predictive validity of the FDTVP, especially on its relationship to reading and IQ. Frostig (1963) maintains that the FDTVP is highly correlated with reading in the early ages, but the low reliability and relative homogeneity of the scores suggests that the predictive validity of the FDTVP would be very low for populations similar to the sample in this study.

The findings of this study on subscale IV for interaction effects of sex of examiners, sex of subject, and administration mode warrant the need for further research on this subscale.

BIBLIOGRAPHY

BIBLIOGRAPHY

ARTICLES

Alden, P., and Benton, A.L. "Relationship of Sex of Examiner to Incidence of Rorschach Responses with Sexual Content." Journal of Projective Techniques, 1951, 15, pp. 230-234.

Allen, R.F. "Factor Analysis of the Developmental Test of Visual Perception Performance of Educable Mental Retardates." Perceptual and Motor Skills, 1968, pp. 257-258.

Anderson, Betty. "Classroom Diagnosis of Reading Readiness Factors." Journal of Learning Disabilities, May, 1970, 3 (5), pp. 160-163.

Barnard, P. "Interaction Effects Among Certain Experimenter and Subject Characteristics on a Projective Test." Journal of Consulting and Clinical Psychology, 1968, 32, pp. 514-521.

Barratt, E.S. "The Space-Visualization Factors Related to Temperament Traits." Journal of Psychology, 1955, XXXIX, pp. 279-287.

Baughman, E.E. "Rorschach Scores as a Function of Examiner Differences." Journal of Projective Techniques, 1951, 15, pp. 243-249.

Bauer, David. "The effect of Instruction, Anxiety, and Locus of Control on Intelligence Test Scores." Measurement and Evaluation in Guidance, April, 1975, 8 (1).

Becker, P., and Sabatino, D. "Frostig Revisited." Journal of Learning Disabilities, 1973, 6 (3), pp. 180-184.

Becker, W.C. "The Matching of Behavior Rating and Questionnaire Personality Factors." Psychological Bulletin, 1960, 57, pp. 201-212.

Beloff, H. "Two Forms of Social Conformity: Acquiescence and Conventionality." Journal of Abnormal and Social Psychology, 1958, 56, pp. 99-104.

Benney, M., Riesman, D., and Star, S. "Age and Sex in the Interview." American Journal of Sociology, 1956, 62, pp. 143-152.

Bieri, J., Bradburn, W., and Galinsky, M.D. "Sex Differences in Perceptual Behavior." Journal of Personality, 1958, XXVI, pp. 1-12.

Binder, A., McConnell, D., and Sjoholm, N. "Verbal Conditioning as a Function of Experimenter Characteristics." Journal of Abnormal and Social Psychology, 1957, 55, pp. 309-314.

Block, J., and Block, J. "An Interpersonal Experiment on Reactions to Authority." In D.C. McClelland (ed.). Studies in Motivation, New York: Appleton-Century-Crofts, 1955, pp. 537-544.

Bortner, M., and Birch, H. "Perceptual and Perceptual-motor Dissociation in Brain-damaged Patients." Journal of Nervous and Mental Disease, 1960, 130, pp. 49-53.

Boyd, L., and Randle, K. "Factor Analysis of the Frostig DTVP." Journal of Learning Disabilities, 1970, 3, pp. 253-255.

Braithwaite, R.J. "The Frostig Test and Training Program - How Valuable?." Slow Learning Child, July, 1972, 19 (7), pp. 86-89.

Broadhurst, A., and Phillips, C. "Reliability and Validity of the Bender-Gestalt Test on a Sample of British School Children." British Journal of Social Clinical Psychology, 1969, 8, pp. 253-262.

Bryan, Q. "Relative Importance of Intelligence and Visual Perception in Predicting Reading Achievement." California Journal of Educational Research, 1964, 15, pp. 44-48.

Bush, D., and Coward, R. "Sex Differences in the Solution of Achromatic and Chromatic Embedded Figures." Perceptual and Motor Skills, 1974, 39, pp. 1121-1123.

Buss, A., and Durkee, A. "Conditioning of Hostile Verbalizations in a Situation Resembling a Clinical Interview." Journal of Consulting Psychology, 1958, 22, pp. 514-518.

Carlson, E., and Carlson, R. "Male and Female Subjects in Personality Research." Journal of Abnormal and Social Psychology, 1960, 61, pp. 482-483.

Cheek, F. "A Serendipitous Finding: Sex Roles and Schizophrenia." Journal of Abnormal and Social Psychology, 1964, 69, pp. 396-400.

Chissom, B.S. "Canonical Validity of Perceptual-Motor Skills for Predicting an Academic Criterion." Educational and Psychological Measurement, 1972, 33, pp. 1095-1098.

Cieutat, V.J. "Examiner Difference with the Stanford-Binet IQ." Perceptual and Motor Skills, 1965, 20, pp. 317-318.

Cieutat, V.J., and Flick, G.L. "Examiner Differences Among Items on the Stanford-Binet." Psychological Reports, 1966.

Clark, R.A. "The Projective Measurement of Experimentally Induced Levels of Sexual Motivation." Journal of Experimental Psychology, 1952, 44, pp. 391-399.

Cohen, Alan. "Some Learning Disabilities of Socially Disadvantaged Puerto Rican and Negro Children." Academic Therapy Quarterly, 1966, 2 (1).

Corah, N., and Powell, B.J. "A Factor Analytic Study of the Frostig DTVP." Perceptual Motivation Skills, 1963, 3 (1), pp. 59-63.

Cordaro, L., and Ison, J.R. "Psychology of the Scientist 'X'. Observer Bias in Classical Conditioning of the Planarian." Psychological Reports, 1963, 13, pp. 787-789.

Crawley, J.F., Burrow, W.H., and Goodstein, H.A. "An Appraisal of Head Start Participants and Non-participants." Research Report, Office of Economic Opportunity, Starrs, Connecticut: University of Connecticut, 1968.

Cronbach, L.J., Rajaratnam, H., and Gleser, G.C. "Theory of Generalizability: A Liberalization of Reliability Theory." British Journal of Statistical Psychology, 1963, 16 (2), pp. 137-163.

Cruise, D. "Effects of Distraction upon the Performance of Brain Injured and Familial Retardates." American Journal of Mental Deficiency, 1961, 66, pp. 86-92.

Crutchfield, R. "Conformity and Character." American Psychologist, 1955, 10, pp. 191-198.

Curtis, H., and Wolf, E. "The Influence of the Examiner on the Production of Sex Responses on the Rorschach." American Psychologist, 1951, 5, pp. 345-346.

Dickstein, L., and Kephart, J. "Effect of Implicit Examiner Expectancey upon WAIS Performance." Psychological Reports, 1972, 30, pp. 207-212.

Dinoff, Michael. "Subject Awareness of Examiner Influence in a Testing Situation." Journal of Consulting Psychology, 1960, 24 (5), p. 465.

Domrath, Richard. "Constructional Praxis and Visual Perception in School Children." Journal of Consulting and Clinical Psychology, 1968, 32, pp. 186-192.

Donofrio, T. "The Effect of Examiner Variables on Amount of Self-disclosure in TAT Stories." Doctoral Dissertation, 1968.

Dymond, R. "Personality and Empathy." Journal of Consulting Psychology, 1950, 14, pp. 343-350.

Ebel, Robert. "The Relation of Scale Fineness to Grade Accuracy." Journal of Educational Measurement, 1969, 6, pp. 217-221.

Egeland, B., Rice, and Penny. "Interscorer Reliability on the Bender-Gestalt Test and the Visual Retention Test." American Journal of Mental Deficiency, 1967-68, 72, pp. 96-101.

Eister, R. "Thematic Expression of Sexual Conflict under Varying Stimulus Conditions." Journal of Consulting and Clinical Psychology, 1968, 32, pp. 216-220.

Exline, R.V. "Effects of Need for Affiliation, Sex and Sight of Others upon Initial Communication in Problem Solving." Journal of Personality, 1962, 30, pp. 541-556.

Fidel, and Ray. "The Validity of the Revised Objective Perceptual Test in Differentiating Among Nonorganic, Minimally Organic and Grossly Organic Children." Journal of Special Education, 1972, 6 (3), pp. 279-284.

Forrester, B.J., and Klaus, R.A. "The Effect of Race of the Examiner on Intelligence Test Scores of Negro Kindergarten Children." Peabody Papers in Human Development, 1964, 2 (7), pp. 1-7.

Forslund, Morris, and Hull, Ronald. "Teacher Sex and Achievement Among Elementary School Pupils' Education." Journal of Special Education, 1974-75, 95, pp. 87-89.

Frostig, Marianne. "Testing as a Basis for Educational Therapy." Journal of Special Education, 1967; 2 (1), pp. 15-34.

Frostig, M., Lefever, and Whittlesey. "A Developmental Test of Visual Perception for Evaluating Normal and Neurologically Handicapped Children." Perceptual and Motor Skills, 1961, 12, pp. 383-394.

Frostig, M., Lefever, and Whittlesey. "The Marianne Frostig Developmental Test of Visual Perception." Perceptual and Motor Skills (Manual), 1964, 19, pp. 463-499.

Fuller, C.B. "The Minnesota Percepto-Diagnostic Test." Journal of Clinical Psychology, 1969, (Monogram Supplement).

Gardner, R.W., Jackson, D.N., and Messick, S. "Personality Organization in Cognitive Controls and Intellectual Abilities." Psychological Issues, II, 1960.

Garfield, S., Blek, L., and Melker, F. "The Influence of Method of Administration and Sex Differences on Selected Aspects of TAT Stories." Journal of Abnormal Social Psychology, 1951, 43, pp. 338-345.

Garmezy, N., and Rodnick, E. "Premorbid Adjustment and Performance in Schizophrenia: Implications for Interpreting Heterogeneity in Schizophrenia." Journal of Nervous and Mental Disease, 1959, 129, pp. 450-466.

Gewirtz, J.L. "Three Determinants of Attention Seeking in Young Children." Monogram of Social Research in Child Development, 1954, 19 (21).

Gewirtz, J.L., and Baer, D.M. "The Effect of Brief Social Deprivation on Behaviors for a Social Reinforcer." Journal of Abnormal Social Psychology, 1958, 56, pp. 49-56.

Gibby, R. "Examiner Influence on the Rorschach Inquiry." Journal of Consulting Psychology, 1957, 16, pp. 449-455.

Golden, Charles, "Sex Differences in Performance on the Stroop Color and Word Test." Perceptual and Motor Skills, 1974, 39, pp. 106-107.

Goodenough, D.R., and Karp, S.A. "Field Dependence and Intellectual Functioning." Journal of Abnormal and Social Psychology, 1961, 63, pp. 241-246.

Gordon, L.V., and Durea, M. "The Effect of Discouragement on the Revised S-B Scale." Journal of Genetic Psychology, 1948, 73, pp. 201-207.

Goulet, L.R. "Age, School Experience, and the Development of Visual Perceptual Memory." Report, March, 1974.

Gray, S. "Perceived Similarity to Parents and Adjustment." Child Development, 1959, 30, pp. 81-107.

Gross, L. "Effects of Verbal and Nonverbal Reinforcement on the Rorschach." Journal of Consulting Psychology, 1959, 23, pp. 66-68.

Guarnaccia, Vincent. "Comparison of Automated and Standard Administration of the Purdue Pegboard with the Mentally Retarded Adults." Perceptual and Motor Skills, 1975, 40, pp. 371-374.

Guilford, J.P., and Lacey, J.L. (eds.). "Printed Classification Tests" (AAF Aviation Psychology Program, Research Reports, Rep. No. 5). Washington: Government Printing Office, 1947.

Haan, N., and Liuson, W. "Sex Differences in the Eyes of Expert Personality Assessors: Blind Spots." Journal of Personality Assessment, 1973, 37 (5), pp. 486-492.

Hammill, D.D., Colarusso, R.P., and Wiederholt, J.L. "Diagnostic Value of the Frostig Tests: A Factor Analytic Approach." Journal of Special Education, 1970, 4 (7), pp. 279-283.

Hammond, K.R. "Representative Versus Systematic Design in Clinical Psychology." Psychological Bulletin, 1954, 51, pp. 150-159.

Harris, Sandra. "Influence of Subject and Experimenter Sex in Psychological Research." Journal of Consulting and Clinical Psychology, 1971, 37 (2), pp. 291-294.

Harris, S., and Masling, J. "Examiner Sex Subject, and Rorschach Productivity." Journal of Consulting and Clinical Psychology, 1970, 34, pp. 60-63.

Henning, J., and Kornreich, L.B. "A Developmental Study of the Effects of Pretraining on a Perceptual Recognition Task." Child Development, 1971, 42, pp. 2117-2119.

Henry, E., and Rotter, J. "Situational Influences on Rorschach Responses." Journal of Consulting Psychology, 1956, 20, pp. 457-462.

Hills, J.R. "Within-group Correlations and Their Correction for Attenuation." Psychological Bulletin, 1957, 54, pp. 131-133.

Hochberg, J., and McAlister, E. "Quantitative Approach to Figural Goodness." Journal of Experimental Psychology, 1953, 45, pp. 361-364.

Holtzman, Wayne. "The Examiner as a Variable in the Draw-a-Man Test." Journal of Abnormal Social Psychology, 1948, p. 43.

Horner, M.S. "Femininity and Successful Achievement: A Basic Inconsistency." Journal of Personality and Social Psychology, 1975, 19, pp. 114-118.

Howard, J. "The Group Bender-Gestalt Test as a Screening Procedure for the Identification of Children with Lags in Visual Perceptual Development." Journal of School Psychology, 1970, 8, pp. 64-65.

Hutt, G.J. "Consecutive and Adaptive Testing with the Revised Stanford-Binet." Journal of Consulting Psychology, 1947, 11, pp. 93-104.

Jackson, D.N. "Stability in Resistance to Field Forces." Unpublished Doctoral Dissertation, Purdue University, 1955.

Jackson, Douglas, Messick, S., and Myers, C.T. "Evaluation of Group and Individual Forms of Embedded Figures Measures of Field Independence." Educational and Psychological Measurement, 1964, 24 (2), pp. 177-192.

Kagan, J. "The Child's Perception of the Parent." Journal of Abnormal Social Psychology, 1956, 52, pp. 257-258.

Kanfer, F.H. "Verbal Conditioning, Reinforcement Schedules and Experimenter Influence." Psychological Reports, 1958, 4, pp. 443-452.

Keough, B.K., and Smith, C.E. "Group Techniques and Proposed Scoring System for the Bender-Gestalt Test with Children." Journal of Clinical Psychology, 1961, 17, pp. 172-175.

Keough, B.K., and Smith, G.E. "Visual-motor Ability for School Prediction: A Seven-year Study." Perceptual and Motor Skills, 1967, 25, pp. 101-110.

Kintz, B.L., Delprato, D.J., Mettee, D.R., Persons, C.E., and Scharpe, R.H. "The Experimenter Effect." Psychological Bulletin, 1965, 63, pp. 223-232.

Klatskin, E.H. "An Analysis of the Effect of the Test Situation upon the Rorschach Record." Journal of Projective Techniques, 1952, 16, pp. 193-199.

Klein, E., Cicchett, D., Spohn, H. "A Test of the Censure-Deficit Model and its Relation to Premorbidity in the Performance of Schizophrenics." Journal of Abnormal Psychology, 1967, 72, pp. 174-181.

Krasner, L. "Studies of the Conditioning of Verbal Behavior." Psychological Bulletin, 1958, 55, pp. 148-170.

Lantz, B. "Some Dynamic Aspects of Success and Failure." Psychological Monograms, 1958, 59, pp. 6-21.

Lasch, L., Holland, J., and Lasch, R. "A New Group Administration Procedure for the Minnesota Percepto-Diagnostic Test." Psychology in the School, 1974, 11, pp. 403-407.

Le Bovitz, Leon. "Testing Experience and Stanford-Binet Scores." Journal of Educational Measurement, 1963, 3 (3), pp. 229-233.

Leiburt, R.E., and Shark, J.K. "Three FDTVP and Specific Reading Tasks for Kindergarten, First and Second Grade Children." Reading Teacher, 1970, 24, pp. 130-137.

Levy, L. "A Note on Research Methodology used in Testing for Examiner Influence in Clinical Test Performance." Journal of Consulting Psychology, 1956, 20, p. 286.

Lord, Edith. "Experimentally Induced Variations in Rorschach Performances." Psychological Monograms, 1950, 64 (70).

Madden, Theodore. "A Note on the Administration and Scoring of the WISC Mazes Subtest." Psychology in the Schools, 1974, pp. 141-143.

Masling, J. "The Effects of Warm and Cold Interaction on the Administration and Scoring of an Intelligence Test." Journal of Consulting Psychology, 1959, 23, pp. 336-341.

Masling, J. "The Influence of Situational and Interpersonal Variables in Projective Testing." Psychological Bulletin, 1960, 57, pp. 65-85.

Masling, J. "The Effects of Warm and Cold Interaction on the Interpretation of a Projective Protocol." Journal of Projective Techniques, 1957, 21, pp. 377-383.

Masling, J., and Harris, S. "Sexual Aspects of TAT Administration." Journal of Consulting Psychology, 1969, 33, pp. 166-169.

Maslow, P., Frostig, M., Lefever, D.W., and Whittlesey, J.R.B. "The Marianne Frostig Developmental Test of Visual Perception: 1963 Standardization." Perceptual and Motor Skills, Monograph Supplement 2, V. 19, 1964.

McClelland, D., and Watt, N. "Sex-role Alienation in Schizophrenia." Journal of Abnormal Psychology, 1968, 73, pp. 226-239.

McConnel, D., Bender, A., and Sjoholm, N. "Verbal Conditioning as a Function of Experimenter Characteristics." Journal of Abnormal and Social Psychology, 1957, 55, pp. 309-314.

McGuigan, F.J. "The Experimenter: A Neglected Stimulus Object." Psychological Bulletin, 1963, 60, pp. 421-428.

McNemar, Q. "Lost, Our Intelligence. Why?." American Psychology, 1964, 19, pp. 871-882.

Melker, F., and Garfield, S. "The Influence of Method of Administration and Sex Differences on Selected Aspects of TAT Stories." Journal of Abnormal Social Psychology, 1951, 43, pp. 338-345.

Miller, C., and Chansky, N. "Psychologists Scoring of WISC." Psychology in the Schools, 1972, 9, pp. 144-152.

Miller, L. "Reliability of Koppitz Scoring System for the Bender-Gestalt." Journal of Clinical Psychology, 1963, 19, p. 211.

Miller, L.C., Loewenfeld, R., Linder, R., and Turner, J. "Reliability of Koppitz Scoring System for the Bender-Gestalt." Journal of Clinical Psychology, 1963, 19, p. 21f.

Mlodnosky, L.B. "Bender-Gestalt and the Frostig as Predictors of First Grade Reading Achievement Among Economically Deprived Children." Psychology in the Schools, 1972, 9, pp. 25-30.

Morsbach, G., Priori, C.D., and Furnell, J. "Two Aspects of Scorer Reliability in the Bender-Gestalt Test." American Journal of Mental Deficiency, 1971, pp. 90-93.

Newcomer, R., and Hammill, D. "Effects of Motor Handicap upon Bender-Gestalt Test Performance." Exceptional Children, 1973, 40 (1), pp. 13-20.

Ohnmacht, F., and Rosen, C.L. "Perception, Readiness and Reading Achievement." Paper presented at meeting of the International Reading Association, Seattle, April, 1967.

Ohnmacht, F.W., and Olson, A.V. "Canonical Analysis of Reading Readiness Measures and the FDTVP." Educational and Psychological Measurement, 1968, 28, pp. 429-484.

Olson, A.V. "Frostig Developmental Test of Visual Perception as a Predictor of Specific Reading Abilities with Second-grade Children." Elementary English, 1966, 43, pp. 869-872.

Olson, A.V., Simpson, H.D., Rosen, C., Olson, N.H., and Rentz, R.A. "A Multivariate Analysis of First Grade Reading Achievement, Reading Readiness and Intelligence." Technical Reports, University of Georgia, Research and Development Centre, 1968.

Olson, A.V., and Johnson, I. "Structure and Predictive Validity of the FDTVP in Grades One and Three." Journal of Special Education, 1970, 4, pp. 49-52.

Overlade, D.C. "Human Perception as Abstraction Ability." Doctoral Dissertation, Purdue University, 1956.

Page, E. "Teacher Comments and Student Performance: A Seventy-four Classroom Experiment in School Motivation." Journal of Educational Psychology, 1958, 49, pp. 173-181.

Pederson, D., Shinedling, M., and Johnson, D. "Effects of Sex of Examiner and Subject of Children's Quantitative Test Performance." Journal of Personality and Social Psychology, 1968, 10, pp. 251-254.

Pheterson, G.I. "Female Prejudice Against Men." Unpublished Manuscript, Connecticut College, 1969.

Piacente, B., Penner, L., and Cohen, Stephen. "Evaluation of the Performance of Examiners as a Function of their Sex and Competence." Journal of Applied Social Psychology, 1974, 4 (4), pp. 321-329.

Postman, L., and Jarrett, R. "An Experimental Analysis of Learning Without Awareness." American Journal of Psychology, 1952, 65, pp. 244-255.

Rabin, A., Nelson, W., and Clark, M. "Rorschach Content as a Function of Perceptual Experience and Sex of the Examiner." Journal of Clinical Psychology, 1954, 10, pp. 188-190.

Ritter, D., and Sabatino, D. "The Effects of Method of Measurement upon a Child's Performance on Visual Perception Tasks." Journal of School Psychology, 1974, 12 (4), pp. 296-299.

Rosenthal, R. "Experimenter Attributes as Determinants of Subjects' Responses." Journal of Projective Techniques, 1963, 27, pp. 324-331.

Rosenthal, R. "Covert Communication in the Psychological Experiment." Psychological Bulletin, 1967, 67, pp. 356-367.

Rosenthal, R., and Fode, R.L. "Psychology of the Scientist: Three Experiments in Experimenter Bias." Psychological Reports, 1963, 12, pp. 491-511.

Rosenthal, R., Fode, K., Friedman, C.J., and Vikan-Kliner, L.L. "Subjects' Perception of Their Experimenter Under Conditions of Experimenter Bias." Perceptual and Motor Skills, 1960, 11, pp. 325-331.

Rosenthal, R., Persinger, G.H., and Fode, K.L. "Experimenter Bias, Anxiety and Social Desirability." Perceptual and Motor Skills, 1962, 15, pp. 73-74.

Rosenthal, R., Persinger, G., and Fode, K. "The Effect of Early Data Returns on Data Subsequently Obtained by Outcome-biased Experiments." Sociometry, 1963, 26, pp. 487-498.

Rothman, Carole. "Differential Vulnerability of WISC Subtests to Tester Effects." Psychology in the Schools, 1974, 11, pp. 300-302.

Rotter, J.B. "Generalized Expectancy for Internal versus External Control of Reinforcement." Psychological Monographs, 1966, p. 80.

Rubin, S. "A Re-evaluation of Figure-Ground Pathology in Brain Damaged Children." American Journal of Mental Deficiency, 1969, 74, pp. 111-115.

Sabatino, D. "What does the FDTVP Measure?." Exceptional Children, March, 1974, 40, pp. 453-454.

Sanders, R., and Cleveland, S.E. "The Relationship Between Certain Examiner Personality Variables and Subjects' Rorschach Scores." Journal of Projective Techniques, 1953, 17, pp. 34-50.

Sarason, I., and Minard, J. "Interrelationships among Subjects, Experimenter and Situational Variables." Journal of Abnormal and Social Psychology, 1963, pp. 87-91.

Sattler, J.M. "Comments on Cieutat's Examiner Differences with the Stanford-Binet IQ." Perceptual and Motor Skills, 1966, 22, pp. 612-614.

Sattler, J., and Winget, B. "Intelligence Testing Procedures as Affected by Expectancy and IQ." Journal of Clinical Psychology, 1970, 26, pp. 446-448.

Silverman, I., Shulman, A.D., and Wiesenthal, D.L. "The Experimenter as a Source of Variance in Psychological Research: Modeling and Sex Effects." Journal of Personality and Social Psychology, 1972, 21, pp. 219-227.

Silverstein, A.B. "Variance Components in Developmental Test of Visual Perception." Perception and Motor Skills, 1965, 20, pp. 973-976.

Silverstein, A. "Reliability and Abnormality of Differences between FDTVP Subtest Scores." Psychology in the Schools, 1975, 10 (2), pp. 204-206.

Simkins, Lawrence. "Examiner Reinforcement and Situational Variables in a Projective Testing Situation." Journal of Consulting Psychology, 1960, 24 (6), pp. 541-547.

Simon, William. "Expectancy Effects in the Scoring of Vocabulary Items: A Study of Score Bias." Journal of Educational Measurement, 1969, 6 (3), pp. 159-164.

Smith, H., and May, T. "Testing Experience and Stanford-Binet Scores." Journal of Educational Measurement, 1966, 3 (3), pp. 229-235.

Smith, P.A., and Marx, W. "Some Cautions on the Use of the Frostig Test: A Factor Analytic Study." Journal of Learning Disabilities, 1972, 5, pp. 357-362.

Spence, K.W. "Anxiety (drive) Level and Performance in Eyelid Conditioning." Psychological Bulletin, 1964, 61, pp. 129-140.

Sprague, R. "Learning Difficulties of First Grade Children Diagnosed by the Frostig Visual Perceptual Tests: A Factor Analytic Study." Unpublished Ph.D. dissertation, Wayne University, 1963.

Stabler, John. "Measurement of Children's Responses to Authority." Psychological Reports, 1967, 20, pp. 587-590.

Stanton, H. "The Relationship between Teacher's Anxiety Level and the Test Anxiety Level of their Students." Psychology in the Schools, 1974, 11, pp. 360-363.

Steingrueber, H.J. "Handedness as a Function of Test Complexity." Perceptual and Motor Skills, 1975, 40, pp. 263-266.

Stevenson, H.W. "Social Reinforcement with Children as a Function of CA, Sex of Experimenter, and Sex of Subject." Journal of Abnormal Social Psychology, 1961, 63, pp. 147-154.

Stevenson, H.W., and Allen, S. "Adult Performance as a Function of Sex of Experimenter and Sex of Subject." Journal of Abnormal and Social Psychology, 1964, 68, pp. 214-216.

Stevenson, H., Keen, R., and Knights, P. "Parents and Strangers as Reinforcing Agents for Children's Performance," Journal of Abnormal and Social Psychology, 1963, 67, pp. 183-186.

Strauss, M. "Examiner Expectancy: Effects on Rorschach Experience Balance." Journal of Consulting and Clinical Psychology, 1968, 32, pp. 125-129.

Stroop, J.R. "Studies of Interference in Serial Verbal Reactions," Journal of Experimental Psychology, 1955, 18, pp. 643-662.

Synder, R., and Synder, P. "Maturational Changes in Visual Motor Perception: An Item Analysis of Bender-Gestalt Errors from Ages 6 to 11." Perceptual and Motor Skills, 1974, 38, pp. 51-57.

Tauber, R. "Identification of Potential Learning Disabilities." Academic Therapy Quarterly, 1966, 2, pp. 116-123.

Thurston, L.L. "A Factorial Study of Perception." Psychometric Monograph, 1944, No. 4.

Tomlinson, J. "Situational and Personality Correlates of Predictive Accuracy." Journal of Consulting Psychology, 1967, 31, pp. 19-22.

Tuddenham, R. "The Influence of a Distorted Group Norm upon Individual Judgment." Journal of Psychology, 1958, 46, pp. 227-241.

Vernor, Allen, and Feldman, Robert. "Tutor Attributions and Attitudes as a Function of Tutee Performance." Journal of Applied Social Psychology, 1974, 4 (4), pp. 311-320.

Ward, J. "Factor Structure of the FDTVP." British Journal of Educational Psychology, 1970, 40, pp. 65-67.

Werner, G. "The Bender-Gestalt Test as a Predictor of Minimal Neurologic Deficit in Children from Eight to Ten Years of Age." Journal of Nervous Disorders, 1966, 143, pp. 275-280.

Werner, P., and Block, J. "Sex Differences in the Eyes of Expert Personality Assessors: Unwarranted Conclusions." Journal of Personality Assessment, 1973, 39 (2), pp. 110-113.

Werner, H., and Strauss, A. "Pathology of Figure Background Relation in the Child." Journal of Abnormal and Social Psychology, 1941, pp. 236-248.

Wickes, T.A. "Examiner Influence in a Testing Situation." Journal of Consulting Psychology, 1956, 20, pp. 23-26.

Wieckert, Robert. "Experimenter Influence in Psychological Testing. Religious versus Layman." Master's Thesis in Education, University of California, 1967.

Wiederholt, J.L. "The Predictive Validity of Frostig's Constructs as Measured by the FDTVP." Dissertations Abstracts International, 1973, 33, p. 1556A.

Winkel, G.H., and Sarason, I.G. "Subject, Experimenter, and Situational Variables in Research on Anxiety." Journal of Abnormal and Social Psychology, 1964, 68, pp. 601-608.

Witkin, H.A. "Individual Differences in Ease of Perception of Embedded Figures." Journal of Personality, 1950, 19, pp. 1-15.

Witmer, J. "The Effects of Verbal Approval and Disapproval upon the Performance of Third and Fourth Grade Children on the WISC." Journal of School Psychology, 1971, 9, pp. 347-356.

Witmer, J., Bornstein, A., and Dunham, R. "The Effects of Verbal Approval and Disapproval upon the Performance of Third and Fourth Grade Children on Four Subtests of the WISC." Journal of School Psychology, 1971, 9, pp. 347-356.

Zaporozkets, A.V. "The Development of Perception in the Preschool Child." Monograph for the Society for Research in Child Development, 1965, 30, pp. 82-101.

Zimmerman, D.W., Williams, R.H., and Burkheimer, G.J. "Dependence of Test Reliability upon Heterogeneity of Individual and Group Score Distributions." Educational and Psychological Measurement, 1968, 28, pp. 41-46.

BOOKS

Anastasi, Anne (ed.). Individual Differences. New York: Wiley, 1965.

Anastasi, Anne. Psychological Testing, 3d. ed. New York: McMillan, 1968.

Brunswick, E. Perception and the Representative Design of Psychological Experiments. Berkeley and Los Angeles: University of California Press, 1956.

Cronbach, L.J. Essentials of Psychological Testing, 3d. ed. New York: Harper and Row Publishers, 1962.

Glasser, A., and Zimmerman, I. Clinical Interpretation of the WISC. New York: Grune and Stratton, 1967.

Goldstein, K. Language and Language Disturbances. New York: Grune and Stratton, 1948.

Guilford, J.P., and Zimmerman, W.S. The Guilford-Zimmerman Aptitude Survey, (2d. ed.). Beverly Hills, California: Sheridan Supply Company, 1956.

Hutt, G.L. The Clinical Use of the Revised Bender-Gestalt Test. New York: Grune and Stratton, 1960.

Hyman, H., Cobb, W., Feldman, J., Hart, C., and Stember, C. Interviewing in Social Research. Chicago: University of Chicago Press, 1954.

Kerlinger, F. Foundations of Behavioral Research. London: Rinehart and Winston, Inc., 1965.

Koppitz, E. The Bender-Gestalt Test for Young Children. New York: Grune and Stratton, Inc., 1964.

Lewin, K. A Dynamic Theory of Personality. New York: McGraw-Hill, 1935.

Linquist, E. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin, 1953.

McGuigan, F.J. Experimental Psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1960.

Messick, S., and Kogan, N. Categorizing Styles and Cognitive Structure. Princetown, New Jersey: Educational Testing Service, Research Bulletin, 1964.

Montessori, M. The Montessori Method. New York: Schocken, 1964.

Pascal, G., and Suttel, B. The Bender-Gestalt Test. New York: Grune and Stratton, 1951.

Potter, M.C. "On Perceptual Recognition." In J.S. Bruner (ed.) Studies in Cognitive Growth. New York: Wiley, 1966.

Rosenthal, R. Experimenter Effects in Behavioral Research. New York: Appleton-Century-Crofts, 1966.

Sarason, S. Psychological Problems in Mental Deficiency. New York: Harper and Row, 1953.

Sarason, S. Anxiety in Elementary School Children. New York: Wiley, 1960.

Seigel, S. Nonparametric Statistics. New York: McGraw-Hill, 1956.

Stanley, J.C., and Hopkins, K. Educational and Psychological Measurement and Evaluation (5th ed.). Englewood Cliffs: Prentice-Hall, Inc., 1972.

Thorndike, R.L. Reliability. Princetown, New Jersey: Educational Testing Service, 1964.

Thorndike, R.L. Measurement and Evaluation in Psychology and Education (3d. ed.). New York: Wiley, 1968.

Walker, H.M. Statistical Inference. New York: Holt, 1953.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Witkin, H.A., Dyk, R.B. Faterson, H.F., Goodenough, D.R., and Karp, S.A. Psychological Differentiation. New York: John Wiley and Sons, Inc., 1962.