1-1964

An Experimental Attempt To Assess Some Effects Of Scholastic Test Score Interpretation On The Self Acceptance And Social Adjustment Of High School Sophomores

Gerald D. Kranzler

AN EXPERIMENTAL ATTEMPT TO ASSESS SOME EFFECTS OF SCHOLASTIC TEST SCORE INTERPRETATION ON THE SELF ACCEPTANCE AND SOCIAL ADJUSTMENT OF HIGH SCHOOL SOPHOMORES

by

Gerald D. Kranzler

B.S. in Chemistry, Jamestown College 1956

M.Ed. in Guidance, University of North Dakota 1959

A Dissertation

Submitted to the Faculty

of the

Graduate School

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Education

Grand Forks, North Dakota

January
1964

This dissertation submitted by Gerald D. Kranzler in partial fulfillment of the requirements for the Degree of Doctor of Education in the University of North Dakota, is hereby approved by the Committee under whom the work was done.

Chairman

Dean of the Graduate School

ACKNOWLEDGMENT

The author wishes to express his appreciation to his teachers in the Department of Counseling and Guidance, the Department of Psychology, and the Department of Education for the encouragement, educational preparation, and experienced advice each one has contributed to his total educational experience.

Special expression of appreciation is due the members of the author's committee for their methodological suggestions, their constructive comments, and their labor and time spent in bringing this work to its final form.

The author wishes to express his deep gratitude to Dr. Paul Munger for his interest, wise counsel, and encouragement; and to Mr. Martin Morstad for his time and effort expended in carrying this experiment out at Grand Forks Central High School.

Above all, a debt of gratitude will always remain to my wife for her patience, understanding, and support, and for the personal sacrifices she cheerfully made to advance my own ambitions and goals.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF ILLUSTRATIONS

CHAPTER I

NATURE OF THE PROBLEM

Introduction

There is a large body of literature available dealing with the topic of psychological testing. Most of this literature deals with facts about the tests themselves--facts such as their validity and reliability. Recently, however, there has been increasing concern with the processes involved in the use of tests with individuals (Goldman, 1961). Such concern seems appropriate when it is noted that over 122 million test booklets and answer sheets are being sold to schools each school year (Traxler, 1959), and that the practice of reporting and interpreting test scores to students is widespread at the high school level at the present time (Hartman, 1963).

The practice of reporting and interpreting standardized test scores to students exists even though there is almost no evidence concerning the effects that this practice may have on the personal or social adjustment of students. The articles that have been written on the subject are based primarily on the "clinical judgment" of those who are, or have been, engaged in counseling with school age youth, and what is said by these people is

often contradictory. For example, Fenton (1924) states, somewhat quaintly, that " . . . scores do penetrate the students' self sentiments. . . . In fact, an intelligence test score may well deal a severe blow in the solar plexus of the ego" (p. 107). Darley (1947) adds that tests " . . . may play part in the most wanton destruction" (p. 21).

On the other hand, Rothney (1952) states that " . . . telling sophomore, junior, and senior high school students their test scores and interpreting these scores to them during counseling interviews seems not to cause significant negative nor disturbing reactions. . . . The interpretation of test scores to counselees . . . is a valuable procedure in counseling of high school youth" (p. 322).

Since at least two of these men (Darley and Rothney) can be considered authorities in the fields of counseling and guidance, their opinions must be respected. The contradictory nature of their opinions, however, suggests the need for an attempt to assess empirically some of the effects that the practice of reporting and interpreting test scores to students may have on the students' personal and social adjustment.

Approach to the Problem

The subjects of this study were assigned at random to one of two treatment conditions: (1) a test interpretation condition, and (2) a control condition. Subjects

assigned to the test interpretation condition were given a report and interpretation of their scholastic aptitude test scores, while subjects assigned to the control condition were not given a report and interpretation of such scores.

After the test scores had been reported and interpreted to all subjects assigned to the test interpretation condition, the criterion measures employed in this study were then administered to the subjects who had been assigned to both treatment conditions.

When these two groups of subjects were compared with respect to their responses on the criterion measures, it was assumed that any statistically significant differences obtained between the two groups could logically be attributed to the differences in treatment conditions.

The criterion measures employed in this study purported to measure the following traits or characteristics: (1) self-acceptance, (2) social adjustment, and (3) ability to estimate scholastic aptitude test scores. These measures are described in detail in Chapter III.

Hypotheses Tested

The null hypotheses tested in this study were:

1. There is no difference in measured self-acceptance between the groups assigned to the two treatment conditions.

2. There is no difference in measured social adjustment between the groups assigned to the two treatment conditions.

3. There is no difference in measured ability to estimate scholastic aptitude test scores between the groups assigned to the two treatment conditions.

4. There is no difference in measured ability to estimate scholastic aptitude test scores between those subjects whose scholastic aptitude test scores are above the median (local norms) and those subjects whose scores are below the median.

It was decided that none of the null hypotheses listed above would be rejected unless the obtained difference was significant at the .05 level or beyond.

Delimitations of the Study

The subjects used in this study were members of the sophomore class at Central High School, Grand Forks, North Dakota. The primary reason for the selection of these subjects was that they had been given a test of scholastic aptitude, but had not been given a report and interpretation of the results of that test.

A secondary consideration was the facility with which research of this kind could be integrated into the regular school program.

Consequently, only those sophomores who met the following criteria were selected as subjects of this study: (1) they had been given a test of scholastic aptitude during the previous year, (2) they had not been given a report and interpretation of the results of that test,

and (3) they had at least one study hall each week.

Scores reported to the subjects in the test interpretation condition were those earned by them on the School and College Ability Test, and these scores were reported and interpreted to them in small groups (rather than individually or in large groups).

Both the measure of self-acceptance and the measure of social adjustment were based on the Index of Adjustment and Values.

The measure of ability to estimate scholastic aptitude test scores was constructed by the present investigator.

Definition of Terms

Many controversies end with the discovery that no real disagreement was involved, only a "semantic confusion." With this problem of ambiguity in mind, an attempt was made to give as concise and clear a meaning as possible for each term within the present study which the writer felt demanded clarification and definition. Wherever appropriate, terms were given both a literary and an operational definition. The literary definition was taken from A Comprehensive Dictionary of Psychological and Psychoanalytical Terms (English and English, 1958), whereas the operational definition specified the operations of measurement involved.

Self-Acceptance. Literary definition: "An attitude

toward one's personal qualities that finds them of unique worth" (p. 485).

Operational definition: In this study self-acceptance is operationally defined as the trait or characteristic which is measured by the self-acceptance scale of the Index of Adjustment and Values.

Social Adjustment. Literary definition: "A person's harmonious relationship with his social environment" (p. 14).

Operational definition: In this study social adjustment is operationally defined as that trait or characteristic which is measured by the social adjustment scale of the Index of Adjustment and Values.

Scholastic Aptitude. Literary definition: "The personal characteristics that make likely a given degree of success in academic pursuits" (p. 40).

Operational definition: In this study scholastic aptitude is operationally defined as that trait or characteristic which is measured by the School and College Ability Test.

Ability to Estimate Scholastic Aptitude Test Scores. This phrase has no widely accepted literary definition. In this study ability to estimate scholastic aptitude test scores was operationally defined as that trait or characteristic measured by an instrument constructed by the present investigator and labeled the Accuracy Index. The Accuracy Index is described in detail in Chapter III.

Treatment Condition. Literary definition: "The environmental stimuli that are systematically varied and brought to bear upon the subject of an experiment or test; the experimental or independent variable" (p. 564).

Need for the Study

A review of the research related to two of the hypotheses of this study--the hypotheses concerning the effects of scholastic test score interpretation on the self-acceptance and social adjustment of students--seems to indicate that very little empirical evidence concerning these problems is available. Furthermore, the statements of authorities in the fields of counseling and guidance concerning these problems are somewhat contradictory.

Many contemporary personality theorists and mental health workers consider both self-acceptance and social adjustment essential to mental health. If these characteristics are, in fact, correlated with mental health, then it seems desirable that a study be made which attempts to empirically assess the effects that a widespread practice--reporting and interpreting scholastic aptitude test scores to students--may have on the self-acceptance and social adjustment of students who have scholastic aptitude test scores reported and interpreted to them.

A review of the research related to the remaining two hypotheses of this study--the hypotheses concerning the ability to estimate scholastic aptitude test scores--

seems to indicate that most of these studies have been poorly designed and that the results of these studies have been largely inconclusive.

A study which replicates the basic ideas of some of these studies but which employs a different research design, different subjects, different research instruments, and so forth, seems desirable.

CHAPTER II

REVIEW OF RELATED RESEARCH

Review

Published studies concerned with reporting and interpreting psychological test scores to subjects have employed several types of research design. The studies reviewed below were classified according to the type of research design employed.

Criterion measures administered after tests interpreted.--In the simplest of designs employed, subjects who had test results interpreted to them were later followed up to see how well they remembered the test scores which were interpreted to them.

Froehlich and Moser (1954), in a study of this type, interpreted Differential Aptitude Test scores to a group of 150 ninth grade students in the schools of Pittsburg, California. Fifteen months after having received an interpretation of their scores, the subjects were asked to recall them.

The main findings of this study were: (1) a large number of the subjects reported their scores inaccurately, (2) the accuracy of the reported scores varied between DAT subtests; correlations between actual percentile ranks on

the eight subtests and recalled percentile ranks ranged from .41 to .57, (3) highest scores were reported more accurately than second-highest scores, (4) there was a positive correlation between measured aptitude and recall of test scores, and (5) the regression phenomenon appeared to be operating, that is, low-scoring students tended to remember their scores as somewhat higher, whereas high-scoring students tended to remember their scores as somewhat lower.

Robertson (1958) reported a similar study done in connection with an orientation program for prospective college freshmen. After the subjects had participated in a testing and counseling program, they were asked to recall the scores which were reported to them. Fifty per cent of the subjects correctly recalled the fourth of the total distribution in which their scores were located. The subjects who did not recall their scores correctly tended to underestimate their scores. The psychological tests employed in this study were the Differential Aptitude Tests, the Diagnostic Reading Test, and the Strong Vocational Interest Blank.

Parker (1957) completed a study at the Northwestern University Counseling Center in 1956. The chief aim of this study was to measure the ability to recall objective test data immediately after counseling (two minutes after the close of the interview) and one month later. The subjects of his study were 136 veterans of the armed

services. The major findings of his study were that immediate recall was somewhat more accurate than delayed recall; and that accuracy of recall showed no relationship to measured personality traits, intelligence, years of education, age, or vocational aspiration.

The purpose of a dissertation completed at Cornell University by Fernald (1961) was to investigate accuracy of recall of psychological test data after a test interpretation interview. The subjects were 92 male college freshmen. The chief findings in this study were that immediate recall was more accurate than delayed recall; and that no statistically significant relationships were found between accuracy of recall and selected personal characteristics, including intelligence, either under conditions of immediate or delayed recall.

Pre- and post-administration of criterion measures. --A second type of research design employed involved asking subjects to estimate their test scores both prior to, and after, being told what their scores were. The difference in accuracy of self-estimate before and after the test interpretation sessions was, then, attributed to having reported scores to the subjects.

In a study of this type with college students as subjects, Johnson (1953) found that: (1) the accuracy of client self-ratings of their aptitudes and interests increased significantly after counseling interviews which included discussion of test results, and significant

increases persisted over a month's time, (2) the largest increases in accuracy of estimate were for intelligence, followed by interests, and then personality, (3) increases in accuracy of self-ratings seemed to be slightly related to intelligence, education, and emotional stability, and (4) there was a slight negative relationship between accuracy of self-estimate and the total number of tests which had been reported to the subjects.

Robertson (1959) used this same pre- and post-counseling design with prospective college freshmen as subjects. He found that, on various items, 43 to 80 per cent of the subjects changed their self-estimates of ability and interests after tests had been reported to them. The results of a two-year follow-up indicated, however, that these increases in accuracy of self-estimate did not have much stability.

Torrance (1954) did a study in connection with a five-day orientation program for entering college freshmen. The subjects were asked to estimate their scholastic aptitude test scores both before and immediately after an orientation period which included an interpretation of test scores. He reported that students tended to overestimate their scores before the orientation period, but that they became more realistic after the orientation period.

Criterion measures administered to test interpretation and control groups.--A third type of research design employed involved the random assignment of subjects

to test interpretation and control conditions. According to authorities on research design, this design is superior to the other two types of design discussed above because of the provision for adequate control groups.

Berdie (1954) employed this type of research design with a group of college freshmen as subjects. Subjects were asked to estimate various types of psychological test scores both before the beginning of the school year and at the end of a quarter in college. During that time subjects assigned to the test interpretation condition were given counseling which included a report and interpretation of psychological test scores. His findings included: (1) only moderate correlations were obtained between estimated scores and actual scores, indicating rather limited self-knowledge on the part of subjects as to their standing compared with others, (2) male subjects assigned to the test interpretation condition increased significantly more than those assigned to the control condition in their ability to accurately estimate Strong Vocational Interest Blank scores, while there was no significant difference in this respect between female subjects assigned to the two treatment conditions, and (3) no significant improvement in ability to accurately estimate scholastic aptitude or personality variables was found for the subjects in either treatment condition.

Froehlich (1957) employed the same basic type of design with high school students as subjects. He found no

significant increase in accuracy of self-ratings for either the control group or the group which had received a report and interpretation of test scores. He also found no significant relationship between accuracy of self-ratings and any of the following characteristics: age, grade in school, and scholastic aptitude.

Different types of test interpretation compared.--A few studies have been reported which compared group versus individual methods of test interpretation.

Lallas (1956) compared three methods of interpreting achievement tests to subjects. The three methods were: a group (classroom) interpretation, an individual interview, and a group interpretation followed by an individual interview. He compared each of these methods with each other and with a control group. He found that all three of the experimental groups improved their ability to estimate achievement test scores to a significantly greater extent than the control group. The greatest improvements in accuracy of self-estimates were found with individual counseling and group-plus-individual counseling. Somewhat less improvement was shown by those subjects who had only group reporting.

Wright (1963) also compared group versus individual methods of test interpretation. His criteria were: (1) accuracy of self-ratings, (2) feasibility of vocational choice, and (3) client satisfaction with test interpretation sessions. He compared the groups assigned to these

methods with subjects assigned to a control condition. Each group consisted of a random sample of 100 college freshmen. Both experimental groups improved accuracy of self-ratings significantly more than the control group, but there was no significant difference between groups assigned to individual and group methods. With regard to feasibility of vocational choice, neither experimental group differed from the control group, nor did experimental groups differ from each other. Subjects assigned to the individual method of test interpretation gave significantly higher ratings (such as "very effective," "could not be better") of the test interpretation sessions than subjects assigned to the group method did.

Some studies compared methods of test interpretation which varied the amount of client participation allowed or encouraged.

Dressel and Matteson (1950) found that subjects who participated most in test-interpretation sessions gained most in accuracy of self-estimate. They indicated, however, that the conclusions of their study should be considered highly tentative due to the small number of clients and counselors involved and the acknowledged crudity of some of the instruments used.

Lane (1952) completed a doctoral dissertation at Columbia University concerned with comparing directive and client-centered approaches to test interpretation. The subjects of the study were 111 high school students. The

chief content of the interviews was the discussion of aptitude and interest inventory data obtained in a two-day testing program for all students. When accuracy of self-estimate was compared for these two groups of subjects, both one week and one month after counseling, no significant differences between the two groups were found.

Lyle Rogers (1954) compared test-centered and self-evaluative techniques of test interpretation. These two methods differed primarily in the extent to which the counselor attempted to seek client participation in the discussion of test results and in the extent to which non-test data were introduced into the interview. The subjects of this study were college freshmen; 41 subjects were given a test-centered interview and 53 were given a self-evaluation interview. No significant differences in accuracy of self-estimate between these two groups of subjects were found.

Gustad and Tuma (1957) investigated the effects of three methods of test introduction and four methods of test interpretation. Fifty-three male subjects seeking help with vocational choice were used as subjects. The three methods of test introduction were: a non-directive approach, an intermediate approach, and a directive approach. Methods of test interpretation varied with respect to level of specificity in the discussion of test results and the extent to which comparisons were made of the clients' test results and their estimates of these results.

When accuracy of self-estimate was compared, no significant differences were found among the groups.

Holmes (1961) completed a dissertation at Boston University that was concerned with a comparison of four techniques used in presenting test information to college freshmen. Three of the methods involved varying degrees of subject participation in the interpretation of test results, while subjects assigned to the fourth method were mailed their test results. An interesting finding in this study was that subjects who were mailed their test results were able to estimate their test scores more accurately than subjects assigned to any of the other three groups! The other three groups did not differ significantly from each other.

Studies employing measures of self-acceptance or social adjustment as the criterion.--A survey of the literature seems to indicate that no reported attempts have been made to determine the effects that reporting and interpreting psychological test scores to subjects may have on their measured feelings of self-acceptance or social adjustment. The only related study which the writer was able to locate was reported almost four decades ago.

Fenton (1924) sent a questionnaire to a group of subjects approximately a year after a test of intelligence had been interpreted to them. Out of a total of 119 respondents (the total number of subjects to whom the questionnaire was sent was not reported), 69 stated that

knowledge of their test score had no effect on their opinion of themselves, 22 that it had affected them agreeably, and 28 that it had a disagreeable effect. Fenton concluded that knowledge of intelligence test scores did change subjects' opinions of themselves, and that intelligence test scores should not be reported to students because of the disagreeable effects that these test scores seemed to have on some individuals.

CHAPTER III

METHODS AND MATERIALS

The Subjects

The subjects of this study were 154 members of the sophomore class at Central High School, Grand Forks, North Dakota during the school year 1963-64. The primary reason for the selection of these subjects was that they had never had scholastic aptitude test scores reported and interpreted to them and could, therefore, be considered naive in this respect. A secondary consideration was the facility with which a study of this kind could be integrated into the regular school program.

Only those sophomores who met the following criteria were selected as subjects of this study: (1) they had been given a test of scholastic aptitude during the previous year, (2) they had not been given a report and interpretation of the results of that test, and (3) they had at least one study hall per week.

The Research Instruments

The School and College Ability Test (SCAT).--According to its publishers (1957, 1958), the SCAT is a measure of the two school-learned abilities termed "verbal

ability" and "quantitative ability." The test yields three scores: (1) a Verbal score, (2) a Quantitative score, and (3) a Total score. The Total raw score is the sum of the Verbal and Quantitative raw scores.

The SCAT's publishers state that the test was designed to measure the ability to undertake the next higher level of schooling. Because the best predictor of future achievement is past achievement, the test relies heavily on school-developed abilities.

Reported reliabilities of the SCAT scales range from .89 to .95 for grades nine and eleven. These reliabilities can be considered adequate for tests of this type.

At grade levels nine and eleven correlations between SCAT scores and subsequent school grades range from .47 to .61. These results, based on studies in ten schools, seem to indicate that the SCAT is about as valid as most similar tests. Reviews in the Fifth Mental Measurement Yearbook report similar coefficients using grades as the criterion predicted. Fowler (1959) concludes that the SCAT shows promise of being an efficient measure of scholastic aptitude.

The Index of Adjustment and Values (IAV).--According to the author of this instrument, the theory behind it is:

> . . . that a person has information relative to his present self-organization (self-concept is a part of this self-organization) and a view of himself as he wishes to be (concept of his ideal self). A significant portion of his behavior is aimed at bridging this

> gap. Furthermore, his self-satisfaction is directly related to the difference he perceives between his self-concept and his concept of his ideal self. Personal maladjustment exists when the discrepancy between these two concepts is sufficiently large as to cause unhappiness.
>
> Social maladjustment, on the other hand, develops in relationships with people where a person perceives himself as more or less adequate than his peers. Social adjustment is possible where a person believes that he is adequate and has worth, dignity, and integrity and where he believes that other people have similar perceptions of themselves (IAV Manual, pp. 5-6, no publication date listed).

When discussing details of the IAV, an examination of the instrument itself will facilitate communication between writer and reader. A copy of the high school form of this instrument has been included in the Appendix for this purpose.

As can be seen, the IAV consists of two forms: a "Self" form and an "Others" form. Basically, a subject uses the "Self" form to tell three things about himself: (1) how he is in respect to trait words, (2) how he feels about being this sort of person, and (3) how he would like to be in respect to the traits. Each of the ratings is made on a five-point scale as shown in the instructions.

To complete the "Others" form, the subject is asked to think of other people at his grade level--in the case of this study, other tenth graders--and to fill out the form as the average member of his group would fill it out for himself.

Scoring the IAV is begun for both the "Self" and "Others" forms by summing the ratings in each of the

columns. Scores obtained from Column II of the "Self" form constitute an operational definition of the variable termed "self-acceptance" in that a subject who consistently rates himself high in this column indicates that he likes being as he is, whereas, the subject who consistently rates himself low in this column indicates that he dislikes being as he is.

The discrepancy between Columns I and III of the "Self" form constitutes another operational definition of self-acceptance. The subject whose discrepancy between the ratings in these two columns is large is relatively less self-accepting than the subject whose discrepancy is small. This is considered true because the subject whose discrepancy is relatively large indicates that, in respect to the trait words, he would like to be different from what he is (for example, in Column I he might say that he seldom is active, but in Column III he might say that he would like to be active most of the time).

A third type of score which may be obtained from the IAV is the categorical score. The IAV Manual states that:

> One of the most meaningful scores from the Index is the categorical score. . . . These scores place a subject in one of four categories: ++, -+, +- or --. The first of the signs of each pair designates attitude toward self and the second sign shows the relationship of the "Others" Column II score to the attitude toward self score. Thus, a ++ person is one who is more accepting of himself than the average of the standardization population and he believes that the average member of his peer group is at least as

> accepting of himself. A -+ person is less accepting of himself than the average of the standardization population and he believes that the average member of his peer group is at least as accepting of himself. A +- person has above average self acceptance but he believes that other people are less accepting of themselves. And a -- person has below average self acceptance and believes that other members of his peer group are even less self accepting (p. 9).

In explaining the meaning of categorical scores, the IAV Manual implies, and the research evidence reported (see below) indicates, that categorical scores may be considered an operational definition of social adjustment.

The first published reports on the IAV reported research conducted on the adult form of this instrument. It was discovered that this form presented numerous testing problems below the twelfth grade; the most serious problem was that the vocabulary level of this form was too difficult for younger students. Consequently, other forms were developed for the elementary school (grades 3, 4, and 5), junior high school (grades 6, 7, and 8), and high school (grades 9, 10, and 11).

Since the high school form of the IAV was employed in this study, evidence concerning the reliability and validity of this form will be presented below.

Based on a sample of grade 10 students, corrected split-half reliability coefficients ranged from .87 to .97. This evidence seems to indicate that this form of the IAV is an instrument of adequate reliability.

Evidence concerning the content validity of this instrument is based on the method used to obtain stimulus

words. Word lists were selected from words used by students at appropriate grade levels to describe themselves and other people. All words retained in the list for each grade level were item-analyzed so that these lists represented reliable ways in which students describe themselves and other people.

Evidence concerning the concurrent validity of the high school form is based on correlations between this form and the adult form. These two forms were administered to ninth and tenth grade students. Correlations between the two forms ranged from .44 to .88. These correlations can be considered high when it is noted that the vocabulary level of the adult form is difficult for ninth and tenth grade students.

Evidence concerning the validity of the categorical scores is based on two studies. In one study teachers were asked to rate students according to their estimates of their pupils' "adjustments"; in the other study pupils were asked to list the ten students they liked best and the ten they liked least. Statistical analysis indicated that categorical scores were related to both teachers' estimates of "adjustments" and pupils' liking for their peers to a statistically significant degree.

According to the IAV Manual:

> It is obvious from the above description of the evidence concerning the validity of the High School form that it leaves much to be desired. If it can be assumed that this form measures the same variables

> as the Adult form (the relatively high correlation between the two forms indicates that this may be true), then further evidence concerning the validity of this instrument is available. . . . The research done on the Adult Form seems to indicate that it is a valid measure of both personal and social adjustment (p. 10).

The Accuracy Index (AI).--The Accuracy Index is the instrument constructed by the investigator. It is reproduced below:

> Last year some of you took the SCAT test. This test consisted of two parts: (1) your ability to work with words (Verbal Ability), and (2) your ability to work with numbers (Quantitative Ability). How did your scores compare with those of other students in your grade? (If you have not been told your scores, make the best guess you can.)
>
> A. In your ability to deal with words (Verbal Ability):
>
> (1) top one-fourth of your class
> (2) second one-fourth of your class
> (3) third one-fourth of your class
> (4) lowest one-fourth of your class
>
> B. In your ability to work with numbers (Quantitative Ability):
>
> (1) top one-fourth of your class
> (2) second one-fourth of your class
> (3) third one-fourth of your class
> (4) lowest one-fourth of your class

The Accuracy Index is scored by comparing subject estimates of SCAT scores with actual scores earned by them.

Procedure Followed in This Study

Preliminary steps.--At the beginning of the school year 1963-64 permission was granted the investigator to use sophomore students attending Central High School, Grand Forks, North Dakota as subjects of this study. Permission was granted with the stipulation that subjects

would be allowed to participate only during their study-hall periods.

SCAT scores earned by all members of the sophomore class were obtained and local norms based on these scores constructed by the investigator. A list of sophomores who had taken the SCAT during the previous year and who had one or more study halls per week was then obtained (some sophomores did not have any study halls during the week). The name of each of the subjects was placed in the upper-left-hand corner of an index card; in the upper-right-hand corner were placed the student's SCAT scores expressed in quartiles (for example, a student might have had the numbers 1-4-2 placed on his card: this would have indicated that his Verbal score ranked in the top one-fourth of his class, his Quantitative score in the lowest one-fourth, and his Total score in the second one-fourth of his class).

Assignment of subjects to treatment conditions.--To facilitate communication, SCAT Total quartile scores will be referred to below as "Levels." Level 1 will include all subjects whose Total score ranked in the top one-fourth of their class, Level 2 all subjects whose Total score ranked in the second one-fourth of their class, and so forth.

An equal number of males and females in each Level was randomly assigned to one of two treatment conditions: a test interpretation condition or a control condition

(these two treatment conditions are described in detail below).

Table 1 indicates the number of subjects at each Level and of each sex randomly assigned to the two treatment conditions. In Table 1, A_1 represents the test interpretation condition and A_2 represents the control condition.

TABLE 1

NUMBER OF SUBJECTS ASSIGNED TO TREATMENT CONDITIONS ACCORDING TO SEX AND LEVEL

Level	Males		Females		Totals
	A_1	A_2	A_1	A_2	
1	10	10	7	7	34
2	13	13	6	6	38
3	10	10	11	11	42
4	10	10	10	10	40
Totals	43	43	34	34	154

Test interpretation condition described.--The subjects who had been assigned to the test interpretation condition were called from their study halls in small groups and sent to a room which had been assigned to the investigator for this purpose. The groups ranged in size from five to ten students. The median size of the groups was seven.

When each of the groups had assembled, the investigator introduced himself by name. No mention was made of the fact that a study was being made. The subjects were told that they had been assembled for the purpose of informing them of their performance on the SCAT test which they had taken while they were in junior high school.

They were told that the SCAT measured two school-learned abilities that were considered very important to success in school: verbal ability, or their ability to deal with words, which would be important in subjects such as English and history; and quantitative ability, or their ability to deal with numbers, which would be very important in courses such as mathematics and science.

To give the subjects some idea as to why these tests had been administered to them, they were informed that the school used these test results for purposes such as ability grouping, identifying students who were not working up to their capacity, and comparing the abilities of the students of one school with those of another school.

The subjects were told that these test results could be useful not only to schools but to the individuals who had taken the tests as well. The examples of the possible uses to individuals which were given were: getting some idea as to how each of the students' abilities compared with those of classmates, getting an indication as to the grades they would be likely to earn in high school, and helping students to choose an appropriate level of

educational aspiration.

After this brief introduction, an index card was given to each subject in the group on which had been placed his name and his SCAT scores. The subjects' attention was directed to the upper-right-hand corner of the card, where, they were told, they would find three numbers. They were informed that the number to the left represented their performance on the Verbal scale of the SCAT, the number in the middle their performance on the Quantitative scale, and the number to the right represented their Total score (they were told that the Total score was an average of the Verbal and Quantitative scores).

As each of the three SCAT scales was mentioned, the name of each was written on a blackboard. This was done to help the subjects remember what each of the numbers represented.

The subjects were told to imagine that the entire sophomore class consisted of 100 students and that the number of items that had been answered correctly by each student had been arranged from highest to lowest. To help them visualize this, a hypothetical distribution of scores was placed on the blackboard.

They were then shown that when the scores had been so arranged, the distribution of scores could then be divided into four equal parts. They were told that a number could be assigned to each of the four equal parts, and that one way to do this would be to assign the number 1 to

the group of students which had scored the highest, the number 4 to the group which had scored the lowest, the number 2 to the group in the second one-fourth from the top, and the number 3 to the group of students whose scores were in the third one-fourth from the top.

Various combinations of quartile scores were then placed under the scale names which had previously been written on the blackboard. These were used as examples of how students who might have earned these quartile scores compared with their classmates in the abilities measured.

Their attention was directed to their Total score. They were told that, of the three scores, the Total score was the best single predictor of success in high school, and that subjects with high scores on this test could expect to receive better grades in high school than subjects with low scores.

At that point it was emphasized that the SCAT was not able to predict school grades perfectly. Some subjects with high SCAT scores would, possibly, receive low grades in high school; and some subjects with low scores would receive high grades. This was explained to them by pointing out the many factors that could have influenced their test performance either positively or negatively.

The subjects were then told that the results of this test could help them to choose the educational level to which they might aspire. Subjects whose SCAT scores were high and who earned good grades in high school could

seriously consider a four-year college as a possibility. Subjects whose SCAT scores were low should seriously consider education or training other than that offered by a four-year college unless either the results of other scholastic aptitude tests indicated that they had more aptitude than the SCAT indicated, or they received very good grades in high school.

The last portion of the test interpretation meeting was devoted to a discussion of some implications of differences between Verbal and Quantitative scores. Subjects were told that a high score on one scale coupled with a low score on the other suggested two possible courses of action: (1) subjects could take elective courses which were related to their higher ability and, thereby, increase the probability of getting good grades, or (2) they could take elective courses related to their lower ability and, thereby, attempt to improve that ability.

After the SCAT had been reported and interpreted to the subjects, they were dismissed and sent back to their study halls.

Control condition described.--While SCAT test scores were being interpreted to the subjects who had been assigned to the test interpretation condition, the subjects who had been assigned to the control condition remained in their study halls. SCAT test scores were not reported to them, and they were not told that they were subjects of this study.

Administration of the criterion measures.--One week after SCAT test scores had been reported and interpreted to the subjects who had been assigned to the test interpretation condition, the subjects assigned to both treatment conditions were called from their study halls and the criterion measures were administered to them. The criterion measures were administered to all subjects in groups which ranged in size from 13 to 33, with a median size of 21.

When the subjects in each of the groups had assembled, they were told that they were about to be given a different kind of test, different in that there were neither correct nor incorrect answers to the questions. They were told that they were being given these tests to "find out how sophomores in high school react to a test of this type."

A copy of the Index of Adjustment and Values was handed to each of the subjects, and each of the six IAV scales was administered by having the subjects read the directions silently while the investigator read them aloud.

When the subjects in each group had completed the IAV, the Accuracy Index was administered to them. It was emphasized that if they did not know what their SCAT scores were, either because they had never been told their scores or because they had forgotten them, they should still make the best guess that they could.

Scoring the criterion measures.--The Inventory of Adjustment and Values was scored in such a manner that one categorical score and two self-acceptance scores were obtained for each subject.

When scoring the Accuracy Index, subjects were given a "score" of accurate if they were able to estimate their quartile rank in class on both Verbal and Quantitative scales correctly. They were given a "score" of inaccurate if they were not able to estimate their SCAT quartile scores accurately.

Statistical techniques employed.--The data obtained from both of the self-acceptance measures were analyzed for statistical significance by means of three-dimensional analysis of variance designs. Data obtained from measures of social adjustment and ability to estimate scholastic aptitude test scores were analyzed by means of the chi square statistic. (These techniques are described in detail in Chapter IV.)

CHAPTER IV

RESULTS

Test of Hypothesis Number One

Hypothesis number one stated: "There is no difference in measured self-acceptance between the groups assigned to the two treatment conditions." Two separate tests of this null hypothesis were made. The first test employed the IAV Column II of the "Self" form as the criterion measure. The second test employed the arithmetic difference between Columns I and III of the "Self" form of the IAV as the criterion measure.

Column II as the criterion measure.--Scores earned by subjects on this measure of self-acceptance were entered in a triple-entry or three dimensional table. This table is graphically represented in Figure 1. In Figure 1, A_1 represents the test interpretation condition and A_2 represents the control condition.

The data from this table were then analyzed statistically by means of a three-dimensional analysis of variance design. The computational formulas used in this analysis were those presented in Lindquist (1953, p. 226).

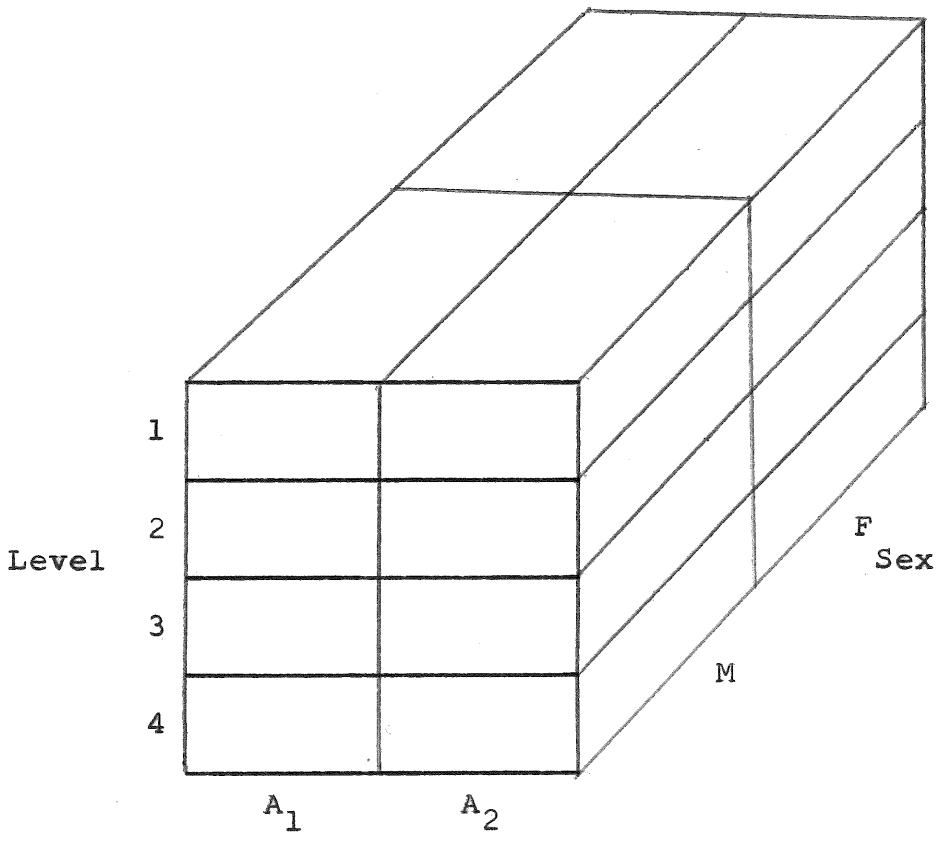

Figure 1.--Graphic Representation of Triple-Entry Table Employed.

Table 2 summarizes the results of this analysis.

TABLE 2

SUMMARY TABLE: TESTING OVERALL SIGNIFICANCE OF DIFFERENCE AMONG GROUPS BY THREE-DIMENSIONAL ANALYSIS OF VARIANCE DESIGN. COLUMN II OF "SELF" FORM AS CRITERION MEASURE

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatments (A)	1	56.16	56.16	.17
Levels (L)	3	1,187.96	395.99	1.20
Sex (S)	1	83.81	83.81	.25
AL Interaction	3	2,088.06	696.02	2.11
AS Interaction	1	683.06	683.15	2.07
LS Interaction	3	987.91	329.30	1.00
ALS Interaction	3	1,294.79	431.60	1.31
Within Cells (W)	138	45,534.01	329.96	. . .
Totals	153	51,915.85	. . .	. . .

Value of F significant at .05 level with 1 and 138 df=3.92
Value of F significant at .05 level with 3 and 138 df=2.68

The interaction effects were tested for significance by: $F = MS_{ALS}/MS_W$, df = 3/138; $F = MS_{AL}/MS_W$, df = 3/138; $F = MS_{AS}/MS_W$, df = 1/138; and $F = MS_{LS}/MS_W$, df = 3/138. When these F ratios were compared with tabled

values of F for the appropriate degrees of freedom (df), it was found that none of the interaction effects reached the pre-determined level of significance (.05). It was, therefore, assumed that MS_w could be used as an appropriate error term when testing the main effects for statistical significance.

The following tests of the main effects were then made: $F = MS_A/MS_w$, df = 1/138; $F = MS_L/MS_w$, df = 3/138; and $F = MS_S/MS_w$, df = 1/138. None of these F ratios reached the pre-determined level of significance (.05).

Because none of the obtained F ratios in this analysis was significant, it was concluded that hypothesis number one could not be rejected when Column II of the "Self" form of the IAV was used as the criterion measure.

Difference between Columns I and II as criterion measure.--Scores earned by subjects on this measure of self-acceptance were entered in a triple-entry table and analyzed for statistical significance by means of a three-dimensional analysis of variance design. The procedure followed was the same as for Column II scores. Table 3 summarizes the results of this analysis.

The value of F which is significant at the .05 level with 3 and 138 degrees of freedom equals 2.68, and the value of F significant at the .05 level with 1 and 138 degrees of freedom equals 3.92. None of the F ratios in Table 4 was computed because it was obvious from mere inspection of the data that none of the computed F ratios

would approach these values.

Because none of these F ratios was statistically significant, it was concluded that it was not possible to reject hypothesis number one when the difference between Columns I and III of the "Self" form of the IAV was used as the criterion measure.

TABLE 3

SUMMARY TABLE: TESTING OVERALL SIGNIFICANCE OF DIFFERENCE AMONG GROUPS BY THREE-DIMENSIONAL ANALYSIS OF VARIANCE DESIGN. DIFFERENCE BETWEEN COLUMNS I AND III OF "SELF" FORM AS CRITERION MEASURE

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatments (A)	1	.06	.06	*
Levels (L)	3	91.09	30.36	*
Sex (S)	1	276.70	276.70	*
AL Interaction	3	563.51	187.84	*
AS Interaction	1	56.51	56.51	*
LS Interaction	3	131.04	43.68	*
ALS Interaction	3	174.67	58.22	*
Within Cells (W)	138	27,939.54	202.46	. . .
Totals	153	29,233.12	. . .	. . .

*not calculated
Value of F significant at .05 level with 1 and 138 df=3.92
Value of F significant at .05 level with 3 and 138 df=2.68

Test of Hypothesis Number Two

Hypothesis number two stated: "There is no difference in measured social adjustment between the groups assigned to the two treatment conditions." The criterion measure employed in the test of this hypothesis was the social adjustment scale of the IAV. The scores earned on this scale were categorical scores (++, -+, +-, and --). According to the IAV Manual, categorical scores range, from least to most desirable, from -- to +- to -+ with ++ as the most desirable.

The data obtained from this criterion measure were analyzed statistically by means of the chi square statistic. To insure an expected frequency within each cell of the chi square table which was large enough to meet the requirements of this statistic, the two most desirable categorical scores (++ and -+) were combined and labeled "good" social adjustment, and the two least desirable categorical scores (+- and --) were combined and labeled "poor" social adjustment.

Preliminary inspection of the data indicated a possible interaction between treatment conditions and Levels. Therefore, two separate tests of hypothesis number two were made, one test for subjects whose SCAT scores were above the median (Levels 1 and 2), and another for subjects whose SCAT scores were below the median (Levels 3 and 4).

Hypothesis tested at Levels 1 and 2.--The frequencies with which "good" and "poor" social adjustment scores were earned by subjects who had been assigned to the two treatment conditions at Levels 1 and 2 were cast into a two-by-two table and analyzed statistically by means of chi square. Table 4 presents the results of this analysis.

TABLE 4

DIFFERENCES AMONG GROUPS IN FREQUENCY OF CATEGORICAL SCORES EARNED AT LEVELS 1 AND 2 ANALYZED BY CHI SQUARE

Social Adjustment Score	Treatment Condition		Chi Square	Probability Less Than
	Test Interpretation	Control		
"good" (++ or -+)	21	17	.94	.50
"poor" (+- or --)	15	19		

Because the obtained chi square did not reach the pre-determined level of significance (.05), it was concluded that hypothesis number two could not be rejected at Levels 1 and 2.

Hypothesis tested at Levels 3 and 4.--The frequencies with which "good" and "poor" social adjustment scores were earned by subjects who had been assigned to the two treatment conditions at Levels 3 and 4 were cast into a

two-by-two table and analyzed by chi square. Table 5 presents the results of this analysis.

TABLE 5

DIFFERENCES AMONG GROUPS IN CATEGORICAL SCORES EARNED AT LEVELS 3 AND 4 ANALYZED BY CHI SQUARE

Social Adjustment Score	Treatment Condition		Chi Square	Probability Less Than
	Test Interpretation	Control		
"good" (++ or -+)	18	29	6.03	.02
"poor" (+- or --)	23	12		

As Table 5 indicates, the frequency with which "good" social adjustment scores were earned was larger in the control condition than in the test interpretation condition, and the frequency with which "poor" social adjustment scores were earned was larger in the test interpretation than in the control condition.

Because the probability associated with the obtained chi square value (less than .02) was less than the pre-determined significance level (.05), it was decided that null hypothesis number two could be rejected for subjects whose SCAT scores were below the median.

The data presented in Table 5 were interpreted as indicating significantly "better" social adjustment for

the control group than for the test interpretation group.

Previous research concerning categorical scores did not conclusively demonstrate a significant difference in the social adjustment of subjects earning intermediate categorical scores (-+ and +-). It was decided, therefore, that an analysis would be made of the frequency with which extreme categorical scores (++ and --) were earned by subjects assigned to the two treatment conditions at Levels 3 and 4. Table 6 presents the results of this analysis.

TABLE 6

DIFFERENCE AMONG GROUPS IN FREQUENCY OF EXTREME CATEGORICAL SCORES AT LEVELS 3 AND 4 ANALYZED BY CHI SQUARE

Categorical Scores	Treatment Condition		Chi Square	Probability Less Than
	Test Interpretation	Control		
++	4	10	4.82	.05
--	11	5		

Because the chi square value obtained in Table 6 was significant at, or beyond, the predetermined level (.05), the data were interpreted as indicating significantly "better" social adjustment for the control group than for the test interpretation group at Levels 3 and 4.

Hypothesis Number Three Tested

Hypothesis number three stated: "There is no difference in measured ability to estimate scholastic aptitude test scores between the groups assigned to the two treatment conditions." This hypothesis was tested by analyzing the frequency with which the two Accuracy Index "scores" (accurate and inaccurate) were earned by the subjects in the two treatment conditions. The frequencies were cast into a two-by-two table and analyzed statistically by means of chi square. Table 7 presents the results of this analysis for all subjects.

TABLE 7

DIFFERENCE IN FREQUENCY OF ACCURACY INDEX SCORES EARNED BY ALL SUBJECTS ANALYZED BY CHI SQUARE

Accuracy Index Scores	Treatment Condition		Chi Square	Probability Less Than
	Test Interpretation	Control		
Accurate	63	14	62.4	.001
Inaccurate	14	63		

Table 8 presents the results of separate analyses at each Level.

TABLE 8

CHI SQUARE ANALYSIS OF DIFFERENCE BETWEEN TREATMENT CONDITIONS IN ACCURACY INDEX SCORES FOR LEVELS 1, 2, 3, AND 4

Level	Accuracy Index Score	Treatment Condition		Chi Square	Probability Less Than
		Test Interpretation	Control		
1	Accurate	16	8	9.06	.01
	Inaccurate	1	9		
2	Accurate	18	3	23.9	.001
	Inaccurate	1	16		
3	Accurate	12	2	5.4	.05
	Inaccurate	9	19		
4	Accurate	17	1	23.9	.001
	Inaccurate	3	19		

Tables 7 and 8 indicate that the differences between the test interpretation and control groups were significant at or beyond the .05 level, both for all subjects and for subjects at each level. Hypothesis number three was, therefore, rejected, and it was concluded that subjects assigned to the test interpretation condition were

able to estimate their scholastic aptitude test scores more accurately than control group subjects.

Hypothesis Number Four Tested

Hypothesis number four stated: "There is no difference in measured ability to estimate scholastic aptitude test scores between those subjects whose scholastic aptitude test scores are above the median (local norms) and those subjects whose scores are below the median." It was decided to make two separate tests of this hypothesis, one for each of the treatment conditions.

Levels 1 and 2 were combined because they consisted of subjects whose SCAT Total scores were above the local median; Levels 3 and 4 were combined because their SCAT Total scores were below the median. The frequencies with which Accuracy Index scores were earned by the subjects in these two groups were cast into two-by-two tables and analyzed by means of chi square.

The results of these analyses are presented in Tables 9 and 10 and indicate that, for both treatment and control conditions, groups of subjects above the median were significantly different from groups below the SCAT median in their ability to estimate scholastic aptitude test scores. It was, therefore, decided that hypothesis number four could be rejected, and it was concluded that subjects whose SCAT Total scores were above the median were able to estimate their scholastic aptitude test

scores more accurately than subjects whose SCAT Total scores were below the median.

TABLE 9

TEST INTERPRETATION CONDITION: CHI SQUARE ANALYSIS OF DIFFERENCES IN ACCURACY INDEX SCORES BETWEEN GROUPS ABOVE AND BELOW SCAT MEDIAN

Accuracy Index Score	Levels 1 and 2	Levels 3 and 4	Chi Square	Probability Less Than
Accurate	34	29	7.24	.01
Inaccurate	2	12		

TABLE 10

CONTROL CONDITION: CHI SQUARE ANALYSIS OF DIFFERENCES IN ACCURACY INDEX SCORES BETWEEN GROUPS ABOVE AND BELOW SCAT MEDIAN

Accuracy Index Score	Levels 1 and 2	Levels 3 and 4	Chi Square	Probability Less Than
Accurate	11	3	6.95	.01
Inaccurate	25	38		

Summary and Discussion

Discussion of the hypothesis concerning self-acceptance.--Analysis of the data indicated that differences in measured self-acceptance between the groups of subjects assigned to the two treatment conditions of this study were not statistically significant. It was felt that there were at least three possible interpretations of these results:

1. Reporting and interpreting scholastic aptitude test scores to subjects had no effect on their feelings of self-acceptance. Goldman (1961), in summarizing studies tangentially related to this one, stated that:

> The total number of hours actually spent by subjects in receiving information about themselves (whether in groups or individually) was very small--probably one to five hours. Perhaps it is unrealistic to expect a great deal of learning to occur in such a short period of time. The time factor is particularly important with self-concept learning; this is not an entirely new subject of study for adolescents and adults. By the time they have reached high school, boys and girls may have reached a practical limit of their self-knowledge, or at least have reached a crystalization of self-concept that is likely to resist change.

Goldman, it seems, would have anticipated the results of the test of hypothesis number one. Because the total amount of time spent in reporting and interpreting test results to each of the subjects of this study amounted to less than an hour, it was, perhaps, unrealistic to expect a change in self-acceptance to occur.

2. A second possible interpretation is that reporting and interpreting scholastic aptitude test scores

to subjects had an effect on their feelings of self-acceptance, but some factor, or combination of factors, in this study resulted in a failure to detect differences between the groups of subjects assigned to the two treatment conditions.

a) Perhaps the IAV does not "really" measure self-acceptance, that is, perhaps the IAV is not valid. The relative paucity of research evidence concerning the validity of the form of the IAV employed in this study does not make it possible to reject this possibility.

b) Perhaps the IAV does measure self-acceptance, but only in a rather gross fashion; that is, the IAV may not be sensitive enough to detect the relatively minor differences in self-acceptance which may have occured in this study.

c) Perhaps changes in self-acceptance occured during, and immediately following, the report and interpretation of test scores, but by the time the criterion measures were administered--one week later--these changes had been dissipated and, thus, no differences between the groups assigned to the two treatment conditions were detected.

3. A third possible interpretation is that reporting and interpreting scholastic aptitude test results had an effect on the subjects' feelings of self-acceptance, but these changes in feelings of self-acceptance were

projected onto others. According to one theory of personality (Rogers, 1951), experiences which are inconsistent with the self-structure of an individual may be perceived as a threat and projected onto others. This possibility will be discussed further in connection with the discussion of hypothesis number 2, below.

Discussion of the hypothesis concerning social adjustment.--Of the groups of subjects assigned to the two treatment conditions, subjects at Levels 3 and 4 differed significantly in measured social adjustment, whereas subjects at Levels 1 and 2 did not.

The possible interpretations of the inability to reject this hypothesis for subjects at Levels 1 and 2 are analagous to those given for hypothesis number 1. That is, failure to reject the hypothesis concerning social adjustment may have been a result of: (1) reporting and interpreting scholastic aptitude tests not having an effect on the social adjustment of subjects, or (2) the social adjustment scale of the IAV not being valid.

It was felt that significant differences between the groups of subjects at Levels 3 and 4 could, logically, be attributed to differences in the treatment conditions, and that it could be concluded that reporting and interpreting test results to subjects in test interpretation groups resulted in "poorer" social adjustment on the part of these subjects.

The writer felt, however, that this conclusion should be considered tentative. When the data based on the IAV were tested for significance, sixteen independent tests were made--seven F ratios for each of the self-acceptance measures and two chi square tests for the social adjustment measure. Of the sixteen tests made, only one significant difference was obtained--the difference in social adjustment between groups assigned to the two treatment conditions at Levels 3 and 4.

It will be recalled that it was decided the null hypotheses of this study would not be rejected unless obtained differences between groups were significant at the .05 level or beyond. It should be noted that when this level of significance is employed, a "Type I" error--rejecting the null hypothesis when it is, in fact, true--will be made approximately five times in every one hundred tests of statistical significance, and one time in every twenty tests of significance. In the present study, when the data derived from the IAV were analyzed, only one out of sixteen tests of significance resulted in a rejection of a null hypothesis. The probability of having committed a Type I error when rejecting that hypothesis was strongly indicated.

Even though there appeared to be a strong possibility that rejection of the null hypothesis concerning social adjustment constituted a Type I error, it was decided that there was also a possibility that a Type I

error had not been made, and that a theoretical explanation of the obtained results should be given.

To aid in this explanation, some analyses of the data were made which were not originally contemplated. Table 11 presents the results of one of these analyses.

TABLE 11

ESTIMATES OF SCAT SCORES MADE BY SUBJECTS AT LEVELS 3 AND 4 ASSIGNED TO CONTROL GROUP

	Estimate of SCAT Scores		
	Under-Estimate	Accurate	Over-Estimate
Number of Subjects	4	3	34

Table 11 shows that control group subjects at Levels 3 and 4 had a distinct tendency to over-estimate their SCAT scores (34 out of 41 subjects, or 83 per cent, over-estimated their SCAT scores). Since subjects were assigned at random to treatment conditions, it can be assumed that subjects at these Levels who had been assigned to the test interpretation condition also had a tendency to over-estimate their scores before actual SCAT scores were reported and interpreted to them.

The definition of categorical scores (see Chapter II) stated that the second sign of each categorical score

referred to the relationship of the "Others" Column II to the attitude toward self (self-acceptance) score. A subject who earned either a +- or a -- categorical score believed that others in his group were less self-accepting than he was, whereas a subject who earned either a -+ or a ++ categorical score believed that others in his group were more self-accepting than he was.

Each of the subjects at Levels 3 and 4 assigned to each of the treatment conditions was categorized according to whether he considered others to be more, or less, self-accepting than himself. Frequencies were cast into a two-by-two table and analyzed statistically by means of chi square. The results of this analysis are presented in Table 12.

TABLE 12

SUBJECTS CLASSIFIED BY TREATMENT CONDITION AND BELIEFS CONCERNING SELF-ACCEPTANCE OF OTHERS. FREQUENCIES ANALYZED BY CHI SQUARE

Others' Self-Acceptance	Treatment Condition		Chi Square	Probability Less Than
	Test Interpretation	Control		
Others More Self-Accepting	18	29	6.03	.02
Others Less Self-Accepting	23	12		

Table 12 indicates that the number of subjects who perceived others as less self-accepting than themselves was larger for the test interpretation than for the control condition, but the number of subjects who perceived others as more self-accepting than themselves was smaller for the test interpretation than for the control condition. The difference in frequencies between the two treatment conditions was significant at the .02 level.

According to Rogers' (1951) theory of personality, experiences which are perceived as inconsistent with the self-structure may be perceived as a threat. Under conditions of threat the self-structure will organize to maintain itself and will build up defenses against threatening experiences by denying them to awareness. According to Hall and Lindzey (1957), "In such a case, the denied feelings may be allowed to express themselves by means of distorted symbolization, for example, by projecting them onto other people."

The data presented in Table 11 seem to indicate that SCAT scores reported and interpreted to subjects may have been perceived as inconsistent with the subjects' self-structure. For subjects at Levels 3 and 4, SCAT scores reported to them were lower than expected and therefore, may have been perceived by these subjects as evidence that they were not as "smart" as they thought they were.

If it can be assumed that a high degree of scholastic aptitude is a cultural ideal (that is, if it can be

assumed that most people in our culture would rather consider themselves "smart" than "dumb"), then telling subjects that they were not as "smart" as they thought they were may have been perceived by them as unpleasant or threatening.

Considering the evidence and theory presented above, it was felt that it was possible to explain some of the results of this study somewhat as follows: SCAT scores reported to the subjects at Levels 3 and 4 (below the median) were lower than expected (see Table 11) and these scores were perceived as unpleasant or threatening (an assumption); the self-structure of these subjects organized to maintain itself (Rogers' theory), which resulted in no significant differences in measured self-acceptance between subjects assigned to the two treatment conditions (Tables 2 and 3); the self-acceptance of subjects was decreased by telling them their SCAT scores, but this change in self-acceptance was projected onto others (see Table 12).

Discussion of the hypothesis concerning ability to estimate scholastic aptitude test scores.--Tests of hypothesis number three seemed to demonstrate rather conclusively that reporting and interpreting test results to subjects increased their ability to estimate their scholastic aptitude test scores. This finding was consistent with most of the previous studies of this problem.

The data presented in Table 7 seem to make the following two conclusions possible:

1. Subjects who were not given a report and interpretation of their scholastic aptitude test scores (subjects assigned to the control condition of this study) were not able to estimate their standing relative to their classmates very accurately. Only 14 of the 77 control group subjects were able to estimate their SCAT Verbal and Quantitative quartile scores accurately.

2. The method employed in this study to report and interpret SCAT scores to subjects was very efficient. Of the 77 subjects who had SCAT scores reported and interpreted to them, 64 (82%) were able to recall their SCAT Verbal and Quantitative quartile scores accurately. These data would seem to support the positions of those who advocate group reporting and interpreting of test results and of those who question the necessity of client participation in this process.

It should be acknowledged, however, that a lack of independence existed between the independent and dependent variables of this portion of the study. Hobbs and Seeman (1955) have discussed this problem, and have concluded that data such as these may represent learning of a very superficial nature, a mere parroting, without any substantial or lasting change in the individual's self-concept or behavior. The tests made in this study of the hypotheses concerning self-acceptance and social adjustment, when

coupled with the conclusions of Hobbs and Seeman, lend support to the idea that the "self-knowledge" gained by those who had tests reported and interpreted to them may have been of a very superficial nature.

Discussion of the hypothesis concerning the relationship between scholastic aptitude and the ability to estimate scholastic aptitude test scores.--The research reviewed in Chapter II led to the conclusion that the evidence concerning the relationship between scholastic aptitude and ability to estimate psychological test scores was rather inconclusive. Tests of hypothesis number four seemed to support the position that the two characteristics were positively correlated--subjects above the median in measured scholastic aptitude were able to estimate their scores more accurately (and to a statistically significant degree) than subjects below the median. This relationship existed for both the groups of subjects assigned to the test interpretation condition and the groups assigned to the control condition.

CHAPTER V

SUMMARY

The Problem

The major purpose of this study was to investigate differences between two treatment conditions: a test interpretation condition and a control condition.

The criteria employed were self-acceptance, social adjustment, and ability to accurately estimate scholastic aptitude test scores. Self-acceptance was operationally defined in two ways: (1) as the trait or characteristic measured by Column II of the "Self" form of the Index of Adjustment and Values (IAV), and (2) as the trait or characteristic measured by the discrepancy between Columns I and III of the "Self" form of the IAV.

Social adjustment was operationally defined as the trait or characteristic measured by the categorical scores of the IAV.

Ability to estimate scholastic aptitude test scores was operationally defined as the trait or characteristic measured by the Accuracy Index, an instrument constructed by the investigator.

A secondary purpose of this study was to investigate the relationship between measured scholastic aptitude

and measured ability to estimate scholastic aptitude scores.

Procedure

The subjects of this study were 154 high school sophomore students who had taken the School and College Ability Test (SCAT) while in junior high school but had not been given an interpretation of the results of that test.

An equal number of subjects of each sex and level of scholastic aptitude was randomly assigned to each of two treatment conditions.

Each of the subjects assigned to the test interpretation condition was given a report and interpretation of his SCAT scores. SCAT scores were reported and interpreted to small groups of subjects, and the scores reported were expressed in quartiles based on local norms.

Subjects assigned to the control condition were not given a report and interpretation of their SCAT scores, and were not informed that they were subjects of this study.

Criterion measures were administered to all subjects of this study one week after SCAT scores had been reported and interpreted to the subjects assigned to the test interpretation condition.

Results

When the data obtained from the two measures of self-acceptance were analyzed by means of a three-dimensional analysis of variance design, it was found that neither the main effects nor the interaction effects were statistically significant. This held true for criterion scores obtained from both measures of self-acceptance. It was felt that the best interpretation of these results was that reporting and interpreting SCAT scores to subjects had no effects on their measured feelings of self-acceptance, and that neither sex nor scholastic aptitude was related to measured self-acceptance.

The data obtained from the measure of social adjustment were analyzed by means of chi square. Preliminary inspection of these data seemed to indicate an interaction between treatment conditions and level of scholastic aptitude. It was decided, therefore, that two separate analyses of these data would be made, one for subjects whose SCAT Total scores were above the median, and another for subjects whose SCAT Total scores were below the median.

For subjects above the median there was no significant difference between the two treatment conditions in the frequency with which criterion scores indicating "good" or "poor" social adjustment were earned.

For subjects below the median, however, the data indicated significantly "poorer" social adjustment for subjects assigned to the test interpretation condition

than for subjects assigned to the control condition. These data were interpreted as indicating that reporting and interpreting SCAT test results had a negative effect on the social adjustment of those subjects who had been assigned to the test interpretation condition. It was felt, however, that this conclusion should be considered tentative because the possibility of having committed a Type I error was strongly indicated. Another interpretation given these data was the possibility of these subjects having projected their own decreased feelings of self-acceptance onto other members of their group.

Analysis of the data obtained from the Accuracy Index seemed to indicate that the method of reporting and interpreting SCAT test scores which was employed in this study was very efficient. Eighty-two per cent of the subjects assigned to the test interpretation condition were able to estimate both their Verbal and Quantitative quartile scores accurately.

The differences in ability to estimate scholastic aptitude test scores between subjects assigned to the two treatment conditions were statistically significant at all levels of scholastic ability, and all differences were in favor of subjects assigned to the test interpretation condition. It was felt, however, that this apparent gain in "self-knowledge" on the part of the subjects assigned to the test interpretation condition may have been of a superficial nature.

A second analysis of the data obtained from the Accuracy Index indicated that subjects whose SCAT scores were above the local median were able to estimate their SCAT scores more accurately than subjects whose SCAT scores were below the local median. The differences between groups of subjects above and below the median were statistically significant within both treatment conditions. Previous research had indicated that the evidence concerning the relationship between scholastic aptitude and accuracy of self-estimate was inconclusive. The data in this study seemed to indicate a positive relationship between these two characteristics.

Conclusions

It is recognized that the results of this study cannot, technically, be generalized beyond the sample employed because the subjects employed in this study were not a random sample from any meaningful population. Furthermore, the generalizations are limited by the particular instruments and methods employed in this study.

With these limitations in mind, the conclusions presented in the following paragraphs were drawn from this study.

Results of this study (and previous studies) seem to indicate that the learning which results from reporting and interpreting psychological test scores to subjects is of a very superficial nature. If learning of this sort is

desired, however, the results of this study also indicate that a group method of test score reporting which involves a minimum of client participation is an efficient means of communicating test score information to clients. If this is true, then it seems that a secondary school counselor might best use his time by first reporting and interpreting test scores to groups of subjects, and then counseling individually with those subjects who request and/or need individual interviews.

The evidence presented in this paper regarding the self-acceptance and social adjustment of subjects who had test scores reported to them seems largely inconclusive. It was not possible to reject stated null hypotheses concerning these characteristics, but, it is felt, this may have been due to inappropriate instruments and/or methodology employed in this study.

Recommendations for Future Research

The writer feels that this study should be replicated at some future date with subjects who score below the median on a test of scholastic aptitude. If the results concerning social adjustment which were found in this study are verified, then implications for counseling would seem to be indicated.

If a replication of this study verifies the results of this study, then, perhaps, individual methods of test reporting should be compared with group methods in order to determine whether or not the results were due to the particular methods employed.

APPENDIX

APPENDIX

"SELF" Instructions for HSIAV

There is a need for each of us to know more about ourselves, but seldom do we have an opportunity to look at ourselves as we are or as we would like to be. On the next page is a list of terms that to a certain degree describe people. Take each term separately and apply it to yourself by completing the following sentence:

I AM A (AN) _____ PERSON.

The first word in the list is jolly, so you would substitute this term in the above sentence. It would read--I am a jolly person.

INSTRUCTIONS FOR COLUMN I (next page)

Then decide HOW MUCH OF THE TIME this statement is like you and rate yourself on a scale from 1 to 5 according to the following key.

1. Seldom, is this like me.
2. Occasionally, this is like me.
3. About half of the time, this is like me.
4. A good deal of the time, this is like me.
5. Most of the time, this is like me.

EXAMPLE: Beside the term JOLLY, number 2 is inserted to indicate that--occasionally I am a jolly person.

INSTRUCTIONS FOR COLUMN II (next page)

Now go to Column II. Use one of the statements given below to tell HOW YOU FEEL about yourself as described in Column I.

1. I very much dislike being as I am in this respect.
2. I dislike being as I am in this respect.
3. I neither dislike being as I am nor like being as I am in this respect.
4. I like being as I am in this respect.
5. I like very much being as I am in this respect.

You will select the number beside the statement that tells how you feel about the way you are and insert the number in Column II.

EXAMPLE: In Column II beside the term JOLLY, number 1 is inserted to indicate that I dislike very much being as I am in respect to the term, jolly. Note that being as I am always refers to the way you described yourself in Column I.

INSTRUCTIONS FOR COLUMN III (next page)

Finally, go to Column III; using the same term, complete the following sentence:

I WOULD LIKE TO BE A (AN) _____ PERSON.

Then decide HOW MUCH OF THE TIME you would like this trait to be characteristic of you and rate yourself on the following five point scale.

1. Seldom, would I like this to be me.
2. Occasionally, I would like this to be me.
3. About half of the time, I would like this to be me.
4. A good deal of the time, I would like this to be me.
5. Most of the time, I would like this to be me.

You will select the number beside the phrase that tells how much of the time you would like to be this kind of person and insert the number in Column III.

EXAMPLE: In Column III beside the term JOLLY, the number 5 is inserted to indicate that most of the time, I would like to be this kind of person.

Start with the word ACTIVE and fill in Columns I, II, and III before going on to the next word. There is no time limit. Be honest with yourself so that your description will be a true measure of how you look at yourself.

"OTHERS" Instructions for HSIAV

We would like to gain a better idea of what you think other people are like. To do this we would like you to: (1) think of other people who are in general like you, for example, other freshmen, sophomore, juniors, or seniors and (2) complete the IAV as you think the average person in this group would complete it for himself. Take each of the 37 words and use it to complete the following sentence for the average person in your reference group:

He is a (an) _____ person.

INSTRUCTIONS FOR COLUMN I (next page)

Then decide how much of the time this statement is like this average person and rate him as he would himself on the following scale:

1. Seldom, is this like he sees himself.
2. Occasionally, this is the way he sees himself.
3. About half of the time, this is the way he sees himself.
4. A good deal of the time, this is the way he sees himself.
5. Most of the time, this is the way he sees himself.

Select the number beside the phrase that tells how much of the time he sees himself this way and insert it

in Column I on the next page.
EXAMPLE: Beside the term JOLLY, number 2 is inserted to indicate that this average person in your reference group sees himself occasionally as a jolly person.

INSTRUCTIONS FOR COLUMN II (next page)

Now go to Column II. Use one of the statements given below to tell how he usually feels about himself as described in Column I.

1. He very much dislikes being as he is in this respect.
2. He dislikes being as he is in this respect.
3. He neither dislikes being as he is nor likes being as he is in this respect.
4. He likes being as he is in this respect.
5. He very much likes being as he is in this respect.

Select the number beside the statement that tells how the average person in your group feels about the way he is and insert the number in Column II.
EXAMPLE: In Column II beside the term JOLLY, the number 1 is inserted to indicate that this person dislikes very much being as he is in respect to the term, jolly. Note that being as "he is" always refers to the way he was described in Column I.

INSTRUCTIONS FOR COLUMN III (next page)

Finally, go on to Column III. Using the same term, complete the following sentence:

He would like to be a (an) _____ person.

Then describe how much of the time this average person in your group would like this trait to be characteristic of him and rate him on the following five point scale:

1. Seldom, he would like this to be him.
2. Occasionally, he would like this to be him.
3. About half of the time, he would like this to be him.
4. A good deal of the time, he would like this to be him.
5. Most of the time, he would like this to be him.

Select the number beside the phrase that tells how much of the time this average person in your group would like to be this kind of person and insert the number in Column III.
EXAMPLE: In Column III beside the term JOLLY, number 5 is inserted to indicate that Most of the time, this average person in your group would like to be this kind of person.

Start with the word ACTIVE and fill in Columns I, II, and III before going on to the next word. There is no time limit.

"SELF"

	I	II	III		I	II	III
a. JOLLY	___	___	___	19. kind	___	___	___
1. active	___	___	___	20. loyal	___	___	___
2. alert	___	___	___	21. neat	___	___	___
3. carefree	___	___	___	22. obedient	___	___	___
4. cheerful	___	___	___	23. patient	___	___	___
5. considerate	___	___	___	24. playful	___	___	___
6. cooperative	___	___	___	25. polite	___	___	___
7. courteous	___	___	___	26. quiet	___	___	___
8. dependable	___	___	___	27. sharing	___	___	___
9. democratic	___	___	___	28. sincere	___	___	___
10. faithful	___	___	___	29. studious	___	___	___
11. friendly	___	___	___	30. sociable	___	___	___
12. generous	___	___	___	31. tactful	___	___	___
13. happy	___	___	___	32. thoughtful	___	___	___
14. helpful	___	___	___	33. thrifty	___	___	___
15. honest	___	___	___	34. trustworthy	___	___	___
16. humorous	___	___	___	35. truthful	___	___	___
17. intelligent	___	___	___	36. understanding	___	___	___
18. interesting	___	___	___	37. unselfish	___	___	___

"OTHERS"

	I	II	III		I	II	III
a. JOLLY	___	___	___	19. kind	___	___	___
1. active	___	___	___	20. loyal	___	___	___
2. alert	___	___	___	21. neat	___	___	___
3. carefree	___	___	___	22. obedient	___	___	___
4. cheerful	___	___	___	23. patient	___	___	___
5. considerate	___	___	___	24. playful	___	___	___
6. cooperative	___	___	___	25. polite	___	___	___
7. courteous	___	___	___	26. quiet	___	___	___
8. dependable	___	___	___	27. sharing	___	___	___
9. democratic	___	___	___	28. sincere	___	___	___
10. faithful	___	___	___	29. studious	___	___	___
11. friendly	___	___	___	30. sociable	___	___	___
12. generous	___	___	___	31. tactful	___	___	___
13. happy	___	___	___	32. thoughtful	___	___	___
14. helpful	___	___	___	33. thrifty	___	___	___
15. honest	___	___	___	34. trustworthy	___	___	___
16. humorous	___	___	___	35. truthful	___	___	___
17. intelligent	___	___	___	36. understanding	___	___	___
18. interesting	___	___	___	37. unselfish	___	___	___

BIBLIOGRAPHY

BIBLIOGRAPHY

Berdie, R. F. "Changes in Self Ratings as a Method of Evaluating Counseling," Journal of Counseling Psychology, I (1954), 49-54.

Bills, R. E. Index of Adjustment and Values Manual. Unpublished manual. University of Alabama.

Cooperative Test Division. Cooperative School and College Ability Tests Technical Report. Princeton, New Jersey: Educational Testing Service, 1957.

Cooperative Test Division. SCAT-STEP Supplement. Princeton, New Jersey: Educational Testing Service, 1958.

Darley, J. G. Testing and Counseling in the High-School Guidance Program. Chicago: Science Research Associates, 1947.

Dressel, P. L., and Matheson, R. W. "The Effect of Client Participation in Test Interpretation," Educational and Psychological Measurement, 10 (1950), 693-706.

English, H. B., and English, A. C. A Comprehensive Dictionary of Psychological and Psychoanalytical Terms. New York: Longmans, Green and Co., 1958.

Fenton, W. "Mental Test Scores and Self-Regard," Educational Administration and Supervision, 10 (1924), 103-07.

Fernald, L. D. "Client Distortion of Psychological Test Data in the Test-Interpretation Interview," Dissertation Abstracts, 22 (1961), 2457.

Froehlich, C P. "A Criterion for Counseling," Psychological Monographs. Vol. 71, No. 15, 1957.

Froehlich, C. P., and Moser, W. E. "Do Counselees Remember Test Scores?" Journal of Counseling Psychology, 1 (1954), 149-52.

Goldman, Leo. Using Tests in Counseling. New York: Appleton-Century-Crofts, Inc., 1961.

Gustad, J. W., and Tuma, A. H. "The Effects of Different Methods of Test Introduction and Interpretion on Client Learning in Counseling," Journal of Counseling Psychology, 4 (1957), 313-17.

Hall, C. S., and Lindzey, G. Theories of Personality. New York: John Wiley & Sons, Inc., 1957.

Hartman, B. J. "Client Behavior Toward Objective Test Data On the Test Interpretation Interview." Unpublished Doctoral dissertation, School of Education, Indiana University, 1963.

Hobbs, N., and Seeman, J. "Counseling," Annual Review of Psychology, 6 (1955), 379-404.

Holmes, J. E. "The Comparison of Four Techniques Used in Presenting Test Information to Freshmen Students," Dissertation Abstracts, 21 (1961), 3379.

Johnson, D. G. "Effect of Vocational Counseling on Self-knowledge," Educational and Psychological Measurement, 13 (1953), 330-38.

Lallas, J. E. "A Comparison of Three Methods of Test Interpretation of the Results of Achievement Tests to Pupils," Dissertation Abstracts, 16 (1956), 1842.

Lane, D. "A Comparison of Two Techniques of Interpreting Test Results to Clients in Vocational Counseling," Dissertation Abstracts, 16 (1952), 1842.

Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. New York: Houghton Mifflin Company, 1953.

Parker, J. J. "A Study of Immediate Recall, Delayed Recall, and Distortion of Objective Test Data Interpreted in Counseling," Dissertation Abstracts, 17 (1957), 305-06.

Robertson, M. H. "A Comparison of Counselor and Student Reports of Counselling Interviews," Journal of Counseling Psychology, 5 (1958), 276-80.

Robertson, M. H. "Results of a Pre-college Testing and Counseling Program," Personnel and Guidance Journal, 37 (1959), 451-54.

Rogers, C. R. Client-centered Therapy; Its Current Practice, Implications, and Theory. Boston: Houghton, 1951.

Rogers, L. A. "A Comparison of Two Kinds of Test Interpretation Interview," Journal of Counseling Psychology, 1 (1954), 224-31.

Rothney, John W. M. "Interpreting Test Scores to Counselees," Occupations, 30 (1952), 320-22.

Torrance, E. P. "Some Practical Uses of a Knowledge of Self-Concepts in Counseling and Guidance," Educational and Psychological Measurement, 14 (1954), 120-27.

Traxler, Arthur E. "Standardized Tests," NEA Journal, 48 (November, 1959), 18-20.

Wright, E. W. "A Comparison of Individual and Multiple Counseling for Test Interpretation Interviews," Journal of Counseling Psychology, 10 (1963), 126-33.