

This dissertation has been 64-9775  
microfilmed exactly as received

BURTON, Robert Leland, 1923-

THE INFLUENCE OF REPORTING CONDITION  
AND THE INFLUENCE OF STUDENT ABILITY  
LEVEL ON THE EFFECTIVENESS OF TEST SCORE  
REPORTING TO TENTH GRADE STUDENTS.

The University of Oklahoma, Ed.D., 1964

Education, psychology

University Microfilms, Inc., Ann Arbor, Michigan

COPYRIGHT BY

ROBERT LELAND BURTON

1964

THE UNIVERSITY OF OKLAHOMA  
GRADUATE COLLEGE

THE INFLUENCE OF REPORTING CONDITION AND THE  
INFLUENCE OF STUDENT ABILITY LEVEL ON THE  
EFFECTIVENESS OF TEST SCORE REPORTING  
TO TENTH GRADE STUDENTS

A DISSERTATION  
SUBMITTED TO THE GRADUATE FACULTY  
in partial fulfillment of the requirements for the  
degree of  
DOCTOR OF EDUCATION

BY  
ROBERT LELAND BURTON  
Norman, Oklahoma  
1964

THE INFLUENCE OF REPORTING CONDITION AND THE  
INFLUENCE OF STUDENT ABILITY LEVEL ON THE  
EFFECTIVENESS OF TEST SCORE REPORTING  
TO TENTH GRADE STUDENTS

BY

*O. R. R. R. R.*  
*O. W. Bridger*  
*Henry Angelino*  
*William W. W.*

DISSERTATION COMMITTEE

## ACKNOWLEDGMENTS

The writer expresses sincere appreciation to the dissertation committee for their support and assistance in this research effort. Special gratitude is extended to Dr. O. J. Rupiper, chairman of the committee and director of the dissertation.

The other members of the committee made a number of significant contributions at various stages during the statistical analysis and the writing process. These members were Dr. Henry Angelino, Dr. William Keown, and Dr. Charles Bridges. Their aid was greatly appreciated.

Several other persons, not members of the committee, also provided valuable assistance. These were Cathryn Batson, Jo Beth Clifton, Margaret Hayes, Betty Keown, Louise Tadlock, Carrifae Mount, Rhio Berthrong, and Frances Olney, all of whom made different but important contributions. Appreciation is extended to the administrative and guidance staffs of the participating school and all other persons who assisted in this effort.

The writer is grateful to all members of the Department of Evaluation and Testing, Extension Division, The University of Oklahoma for their moral support and unlimited

patience. Special acknowledgment is expressed to the Co-operative Test Division of the Educational Testing Service for granting permission to include a copy of the Student Report Form in Appendix B.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS . . . . .	iii
LIST OF TABLES. . . . .	vii
LIST OF FIGURES . . . . .	ix
 Chapter	
I. INTRODUCTION.. . . .	1
Review of the Research on Test Reporting .	8
Related Research Conducted at the Secondary School Level. . . . .	8
Related Research Conducted with College Freshmen. . . . .	13
Other Related Research . . . . .	16
Summary of the Research . . . . .	16
II. THE PROBLEM AND EXPERIMENTAL DESIGN . . . . .	18
Introduction . . . . .	18
Statement of the Problem . . . . .	19
Operational Definitions. . . . .	20
Hypotheses . . . . .	22
Experimental Design. . . . .	24
Instruments Used in the Study. . . . .	26
The Self-Rating Scale . . . . .	26
The Achievement Battery . . . . .	28
The Academic Ability Test . . . . .	30
The Test Profile. . . . .	31
The Supplement to the Test Profile. . .	32
The Sample . . . . .	32

III. COLLECTION OF THE DATA. . . . .	37
Procedures for Administration of the Pretest Questionnaire and the Standardized Tests. . . . .	37
Scoring the Tests. . . . .	37
Recording Test-estimates and Scoring the Pre- test Questionnaires for Congruence. . . . .	38
Preparations for Score Reporting . . . . .	39
Score Reporting Procedures . . . . .	40
Marking the Self-estimate Criterion Instrument	43
Special Marking Instructions Used With the Control Group . . . . .	43
IV. STATISTICAL ANALYSES . . . . .	46
Introduction . . . . .	46
Analysis of Pretest Questionnaire (PT-Q) Data. . . . .	47
Analysis of Control and Reliability Data . . . . .	49
Analysis of the Immediate Recall (IR-Q) Data . . . . .	49
Analysis of the Delayed Recall (DR-Q) Data . . . . .	54
Analysis of Data Obtained from Modifying the Marking Instructions for the Criterion Questionnaire . . . . .	60
Further Analyses and Conclusions . . . . .	65
V. SUMMARY, CONCLUSIONS, AND IMPLICATIONS. . . . .	66
Summary. . . . .	66
Conclusions. . . . .	67
Implications for Counselors. . . . .	70
Recommendations for Further Research . . . . .	71
BIBLIOGRAPHY . . . . .	72
APPENDIXES . . . . .	76
A. Statistical Data . . . . .	76
B. Instruments and Materials. . . . .	102
C. Typescripts. . . . .	117



# LIST OF TABLES

Table	Page
1. Summary of Bartlett's Test of Homogeneity of Variance Chi-Square Values and Analysis of Variance <u>F</u> Values for Groups Matched For Ability Level Using SCAT Raw Scores . . . . .	34
2. Significance of the Difference Between SCAT (T) Raw Score Means by Ability Level. . . . .	35
3. Summary Analysis of Variance for the Pretest Questionnaire Data. . . . .	48
4. Summary Analysis of Variance for the Immediate Recall Data . . . . .	52
5. Significance of the Differences Between Means of Gain for Immediate Recall Data by Ability Level . . . . .	53
6. Summary Analysis of Variance for the Delayed Recall Data . . . . .	56
7. Significance of the Differences Between Means of Gain Scores for Delayed Recall Data by Ability Levels. . . . .	58
8. Correlations Between Three Different Marking Instructions--Control Group Only. . . . .	61
9. Significance of the Difference Between Means of Gain Scores When Three Different Marking Instructions Were Used--Control Group Only. . .	63
10. Cell One, Reporting Condition-Group, Ability Level-Low . . . . .	78
11. Cell Two, Reporting Condition-Group, Ability Level-Middle. . . . .	79
12. Cell Three, Reporting Condition-Group, Ability Level-High. . . . .	80
13. Cell Four, Reporting Condition-Individual, Ability Level-Low . . . . .	81

Table	Page
14. Cell Five, Reporting Condition-Individual, Ability Level-Middle. . . . .	82
15. Cell Six, Reporting Condition-Individual, Ability Level-High. . . . .	83
16. Cell Seven, Reporting Condition-SRF, Ability Level-Low . . . . .	84
17. Cell Eight, Reporting Condition-SRF, Ability Level-Middle. . . . .	85
18. Cell Nine, Reporting Condition-SRF, Ability Level-High. . . . .	86
19. Cell Ten, Reporting Condition-Control, Ability Level-Low . . . . .	87
20. Cell Eleven, Reporting Condition-Control, Ability Level-Middle. . . . .	88
21. Cell Twelve, Reporting Condition-Control, Ability Level-High. . . . .	89
22. Distribution of SCAT (T) Raw Scores--Grade Ten .	90
23. Summary Bartlett's Test of Homogeneity of Variance by Ability Level--SCAT (T) Raw Scores. . . . .	91
24. Summary Analysis of Variance by Ability Level-- SCAT (T) Raw Scores . . . . .	92
25. SCAT (T) Raw Score Means and Variances by Ability Level and Reporting Condition . . . .	93
26. Pretest Questionnaire Means and Variances by Ability Level and Reporting Condition . . . .	94
27. Immediate Recall (IR) Means and Variances of Gain Scores by Reporting Condition and Ability Level . . . . .	95
28. Consolidation of Immediate Recall (IR-Gain) Data by Ability Level . . . . .	96
29. Consolidation of Delayed Recall (DR-Gain) by Ability Level . . . . .	96

Table	Page
30. Delayed Recall (DR) Means and Variances of Gain Scores by Reporting Condition and Ability Level. . . . .	97
31. Means, Variances, and Standard Deviations of Gain Scores for Group Reporting of Test Results to Control Groups Using Three Different Marking Instructions. . . . .	98
32. Intercorrelations Between Five Different Markings of the Criterion Instrument by the Control Group. . . . .	99
33. Significance of the Difference Between Means of Gain Scores When Three Different Marking Instructions Were Used--Control Group Only . .	100
34. Direction and Magnitude of Errors in Self-Estimates by Ability Level and Marking Instruction--Control Group Only. . . . .	101

## LIST OF FIGURES

Figures	Page
1. Twelve cell two-part analysis of variance design for analyzing the influence of reporting conditions, ability levels, and the interaction between the two on the effectiveness of test score reporting with equal observations in each cell. . .	27
2. Immediate Recall Means of Gain Scores Graphed by Reporting Conditions and Ability Levels. . . . .	51
3. Delayed Recall Means of Gain Scores Graphed by Reporting Conditions and Ability Levels. .	55
4. Means of Gain Scores Graphed by Ability Levels and Times Obtained . . . . .	59
5. Means of Gain Scores Graphed by Ability Levels and Marking Instructions . . . . .	62

THE INFLUENCE OF REPORTING CONDITION AND THE  
INFLUENCE OF STUDENT ABILITY LEVEL ON THE  
EFFECTIVENESS OF TEST SCORE REPORTING  
TO TENTH GRADE STUDENTS

CHAPTER I

INTRODUCTION

Since World War I the use of standardized tests in American schools has grown to monumental proportions. Data, contributed by a number of investigators, confirm that approximately one million tests per school day are currently being used in American schools.<sup>1</sup> A recent survey of Oklahoma schools having testing programs indicated that an average of 9.6 standardized tests per pupil were administered during the student's twelve years of attendance.<sup>2</sup>

Womer gives the following summary of the more recent influences that have reinforced the expansion of testing in the schools:

---

<sup>1</sup>Howard B. Lyman, Test Scores and What They Mean (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1963), p. 3.

<sup>2</sup>G.R. Snider and H. B. Hunnicut, "Testing in Oklahoma Schools," Oklahoma Teacher, Vol. XLII (1961), pp. 12-17.

We are in a boom period of standardized testing in elementary and secondary schools. Millions of tests are administered each year to pupils at all grade levels--achievement tests, mental ability tests, aptitude tests and interest inventories, as well as several other types of tests and inventories. Some of these tests are given for college scholarship purposes and some for college admissions purposes. Title V of the National Defense Education Act has stimulated, and in some instances required, additions to testing programs at the secondary level. In general, however, these external influences account for a relatively small percentage of the total standardized testing undertaken by a school system.

There are at least two factors which have had a greater impact upon the amount of testing done in the schools than NDEA or college requirements. First, there has been and continues to be a natural growth of standardized testing in all grade levels. Second, the rapid growth of the guidance movement has meant a corresponding rapid growth in testing. This latter influence may well be the most influential one operating, for in many schools the testing program is developed by and operated by guidance personnel.<sup>3</sup>

The results obtained from these testing efforts affect virtually all facets of the school program. Among the more significant uses of the results are the following: curriculum evaluation, grouping of students according to academic ability, diagnosing learning difficulties, and counseling students regarding educational and vocational decisions. Many writers in the field of guidance emphasize that one of the major uses of test results is that of increasing student self-understanding, which implies that the results of tests should be reported to the student.

---

<sup>3</sup>Frank B. Womer, "Testing Programs-Misconception, Misuse, Overuse," Michigan Journal of Secondary Education, Spring (1961), p. 153.

Lundy and Shertzer, in a paper delivered at the 1963 American Personnel and Guidance Association National Convention, made the following statement:

Counselors are charged with the responsibility of assisting a counselee in developing a perceptive understanding of his unique strengths and weaknesses. It is often asserted that the counselee with this insight will make more appropriate choices and decisions. If a counselor is to meet this responsibility he must make use of many resources. One valuable available resource is the effective and appropriate use of test results. A most telling current criticism against guidance personnel is that tests are administered, scores are recorded and forgotten, and the counselor then moves on to the administration of the next testing battery.<sup>4</sup>

Other authors also emphasize the value of standardized test results for increasing self-understanding by allowing the student to compare his performance against a distribution of scores from a definitive population. Yet, when a secondary school counselor seeks to determine how he might effectively carry out this responsibility, he is faced with a paucity of research evidence regarding this problem. In fact, there is a growing controversy over the relative merits, and even the ethicality of some of the current score reporting practices. On the one hand the counselor is urged by the major textbook writers to discuss test results only within the context of the counseling relationship, while at the same time, some of the more reputable and responsible

---

<sup>4</sup>Charles T. Lundy and Bruce Shertzer, "The Relationship of D.A.T. Scores to High School Marks," Paper read before the American Personnel and Guidance Association National Convention, Boston, Massachusetts, April 8, 1963.

test publishers are providing so called "self-interpreting" profiles designed to interpret the meaning of the student's scores from a particular test battery.

Some test users would consider these publications as mere aids to interpretation and contend that their proper use is still restricted to the structure of the counseling interview. Chauncey and Dobbin, however, give support to their use outside the counseling relationship when they state:

To aid in parent acceptance, a few publishers have prepared materials especially to be sent home with students after an important testing, explaining the school's uses of the test information and providing general interpretations of the individual student's scores. The sharing of test score information with students and their parents is standard practice in most testing; a good test should have instructions or materials to make it easy for the school to do it.<sup>5</sup>

Glanz in his recent book, Groups in Guidance, points out that effective self-interpretation is a relatively new concept in most schools. He recognizes that students at the secondary level have long been told their percentile ranks, quartile standings, or other derived scores from their performance on standardized tests. He further indicates that there is a new stress on students' use of test results and refers to several programs which are using test results in group situations to increase student self-understanding.<sup>6</sup>

---

<sup>5</sup>Henry Chauncey and John E. Dobbin, Testing-Its Place in Education Today (New York: Harper and Row, Publishers, 1963), p. 77.

<sup>6</sup>Edward C. Glanz, Groups in Guidance (Boston: Allyn



When test results are to be the focus of the experience, Tyler is strongly opposed to anything but individual counseling or small groups set up especially for group counseling. Her feelings on the subject are well expressed in the following quotation:

It must be recognized by everyone concerned with the program, administrators and teachers as well as counselors, that the reporting of test results is a matter for individual, not group discussion, except in cases where small groups, set up especially for group counseling, have been organized. To save time by giving tests in groups does not weaken a program seriously, but to attempt to make further economies by allowing students to interpret their own scores undermines it completely. It is better to do no testing at all and leave students to judge their capabilities from other kinds of evidence than to open the way for the many kinds of misconceptions and anxieties about intelligence, aptitudes,<sup>7</sup> and interests that testing without counseling creates.<sup>7</sup>

The Ethical Standards of the American Personnel and Guidance Association, Section A, paragraph ten states:

The member shall offer professional services only through the context of a professional relationship. Thus testing, counseling, and other services are not to be provided through the mail by means of newspaper or magazine articles, radio or television programs, or public performances.<sup>8</sup>

The publication of the above paragraph as part of APGA's Ethical Standards did not go unchallenged. In

---

and Bacon, Inc., 1962), pp. 210-211.

<sup>7</sup>Leona E. Tyler, The Work of the Counselor (New York: Appleton-Century-Crofts, Inc., Second Edition, 1961), pp. 129-130.

<sup>8</sup>"Ethical Standards-American Personnel and Guidance Association," The Personnel and Guidance Journal, Vol. XL, No. 2 (October, 1961), p. 207.

referring to Section A, paragraph ten, Page suggested that the statement was premature and unsound. The following is a quotation that appears pertinent to this discussion:

Heretical as it may appear to some, various universities of all sizes give students their test results without one-to-one counseling interviews. Some have meetings where scores are passed out and general comments made to groups of students; others mail forms with enclosed scores and guides to interpretation. Where these practices exist, the beliefs are held that (1) such scores are not injurious; (2) some feedback to students is a good idea; and (3) individual interviews are not, for practical reasons available to all. I am now reluctant to cite institutions doing these things because, obviously, some counselors would regard such practices as unethical! However, I see no persuasive evidence to support their criticism and regard such premature dogmatism as out of character in the guidance profession.

As for test companies: The Educational Testing Service sends directly to each student his own GRE scaled scores and percentile ranks, together with general explanations to help him interpret and use them.<sup>9</sup>

Leo Goldman discusses the reporting of test results only in the counseling setting; however, he reviews several research studies in which other methods of reporting scores were used. He uses the following lead sentence in introducing two studies that found no statistically significant differences between group and individual reporting of scores: "There is encouraging evidence from two of the studies (Lallas, 1956: Wright, 1957) that group reporting techniques may be as effective, or nearly as effective as individual

---

<sup>9</sup>Ellis B. Page, "Counseling With Mass Media," Letter to the Editor, The Personnel and Guidance Journal, Vol. XLI, No. 8 (April, 1963), p. 729.

reports."<sup>10</sup> Goldman's use of the phrase "there is encouraging evidence" would imply that he is maintaining an open mind and reserving judgment regarding the relative merits of various score reporting procedures.

In addition to the conflicting viewpoints on the subject of test score reporting the counselor faces an additional dilemma, the time factor. Wrenn, making reference to an American School Counselor Association (ASCA) study, reported that the median counselor-student ratio is 1 to 412 and that slightly over 10 percent of the counselors operate in schools where the ratio exceeds 1 to over 1000.<sup>11</sup> The school in which the present study was conducted has a counselor-student ratio of approximately 1 to 750. If the counselor could devote 65 percent of his time to student interviews, the reporting of test results on an individual basis would require two and one-half weeks interview time for each 100 students assigned. In the case of this study the counselor would have required almost 19 weeks to complete the task. Under these circumstances it is not surprising that many counselors succumb and use methods of reporting test results to students on other than an individual basis. Some do so, however, with strong doubts and often with

---

<sup>10</sup>Leo Goldman, Using Tests in Counseling (New York: Appleton-Century-Crofts, Inc., 1961), p. 353.

<sup>11</sup>C. Gilbert Wrenn, The Counselor in a Changing World (Washington, D.C.: American Personnel and Guidance Association, 1962), p. 114.

feelings of guilt.

### Review of the Research on Test Reporting

Several contemporary writers have compiled up-to-date reviews of the research on the reporting of test results. The reader is referred to Goldman,<sup>12</sup> Ohlsen,<sup>13</sup> and Lister and Ohlsen,<sup>14</sup> for comprehensive coverage of the subject. For purposes of this study the researches cited will be limited to those most pertinent to score reporting at the secondary school level and which use congruence between student self-estimates and obtained test-estimates as a criterion of self-understanding.

### Related Research Conducted at the Secondary School Level

In a study designed to investigate the accuracy of student recall of test scores following test interpretation, Froelich and Moser administered the Differential Aptitude Tests to 150 ninth grade students. They presented the test results in a series of group meetings at which time each

---

<sup>12</sup>Leo Goldman, op. cit., pp. 345-363.

<sup>13</sup>Merle M. Ohlsen, "Interpretation of Test Scores," The Impact and Improvement of School Testing Programs, National Society for the Study of Education, Sixty-Second Yearbook, Part II (Chicago: The University of Chicago Press, 1963), pp. 256-266.

<sup>14</sup>James L. Lister and Merle M. Ohlsen, The Effect of Orientation to Testing on Motivation for and Outcomes of Test Interpretation, Cooperative Research Project No. 1344, Report based in part upon Mr. Lister's Doctoral Dissertation (Graduate College, University of Illinois, 1962), pp. 11-25.

student drew a copy of his test profile. The students were allowed to keep their profiles and were encouraged to request further explanation if desired. After a fifteen month delay the students were requested to reconstruct their test profiles from memory. The investigators found that a large proportion of the students did not report their scores accurately. Errors in reporting were made by both high and low ranking students; however, a larger proportion of brighter students reported scores accurately than did the less bright.<sup>15</sup>

Adamek analyzed changes in self-perceptions of interests, abilities, and problem areas under two score reporting conditions: (1) test results interpreted by qualified counselors; and (2) test results, in the form of self-interpreting profiles, given to students in small groups. The investigator administered to ninth graders a battery of tests which included the Chicago Primary Mental Abilities, the Kuder Preference Record (Vocational), and Form A of the SRA-Youth Inventory. Adamek concluded:

(a) there were no systematic differential effects by which one type of test interpretation could be concluded to be superior to the other, (b) testing had no differential effects on accuracy of self-perceptions, (c) testing did not induce greater accuracy of self-perceptions, and (d) the use of self-interpreting materials did not elicit more accurate self-perceptions.<sup>16</sup>

---

<sup>15</sup>C.I. Froelich and W.E. Moser, "Do Counselees Remember Test Scores?", Journal of Counseling Psychology, Vol. I (1954), pp. 149-152.

<sup>16</sup>Edward G. Adamek, "The Effects of Testing and Test

Lister and Ohlsen investigated the effects of pre-testing orientation upon student motivation for learning test results and the changes in accuracy of self-estimates resulting from test interpretation. One hundred students in grades five, seven, nine, and eleven received the orientation treatment prior to taking a battery of achievement, intelligence, and interest tests. An equal number in each of the four grades received no orientation (control). All interpretations were made on an individual basis by qualified counselors. The data indicated that (1) the orientation treatment resulted in greater motivation for learning test results within grades seven and nine but not in grades five and eleven; (2) there was no evidence of relationship between orientation and changes in accuracy of self-estimates following test interpretation; (3) significant increases in accuracy of self-estimates were found following test interpretation for all groups at all levels; (4) significant decreases occurred during a two month period following interpretation; and (5) in spite of these decreases, accuracy of self-estimates in the follow-up remained significantly higher than at the time of test administration.<sup>17</sup>

Singer and Stefflre, in studying interests of high

---

Interpretation on Selected Self-perceptions" (Unpublished Doctoral Thesis, University of Chicago, 1961).

<sup>17</sup>James L. Lister and Merle M. Ohlsen, op. cit., pp. 96-97.

school students, obtained self-estimates of interests prior to testing and counseling and again three months following test interpretation. Some statistically significant differences between means and standard deviations were found, indicating less discrepancy between measured and estimated interest scores following counseling. Some sex differences were noted. The authors also suggested that researchers using self-rating techniques should examine the direction of discrepancies between test-estimates and self-estimates as well as the mere size of discrepancy.<sup>18</sup>

A comparison of individual and group methods of reporting test results was conducted by Lallas. He utilized a self-rating scale to evaluate three different methods of test interpretation; individual counseling interview, group interpretation, and group interpretation plus individual interview. He investigated eleventh grade subjects who took the Iowa Tests of Educational Development (ITED). Both before and after taking this battery of tests the subjects estimated their ranks in the various subtests and also their degree of confidence in their estimates. Data obtained following test interpretation indicated the following: (1) the greatest improvements in the accuracy of self-estimates were found with individual counseling and group-plus-individual counseling; (2) less improvement was made by

---

<sup>18</sup>Stanley L. Singer and Buford Stefflre, "Analysis of Self-Estimate in the Evaluation of Counseling," Journal of Counseling Psychology, Vol. I (1954), pp. 252-255.

those receiving group interpretation only; and (3) all three experimental groups showed significant improvement over a control group.<sup>19</sup>

Schulman studied self-understanding among secondary school students. He hypothesized that the learning experiences of the three year period of high school would contribute to a more realistic appraisal of their abilities. Samples of ninth and twelfth-grade students were matched and compared to determine differences in accuracy of self-understanding. The Differential Aptitude Tests (DAT) were used in this investigation. Before the tests were administered, the students rated themselves on each of the subtests of the DAT. This was done after group discussions led by school counselors had provided the students with an understanding of the required task. Both analysis of variance and correlation were used to determine the differences in accuracy of self-report between ninth and twelfth-graders. Schulman concluded from his data that (1) high school seniors have a more realistic understanding of their relative position on the abilities measured by the DAT than do high school freshmen; (2) accuracy of self-report of both groups was positively related to the extent to which the abilities measured corresponded with content areas in the high school curriculum; and (3) during this three year period, continuous

---

<sup>19</sup>John E. Lallas, "A Comparison of Three Methods of Interpretation of the Results of Achievement Tests to Pupils," Dissertation Abstracts, Vol. XVI (1956), p. 1842.



and well defined instruction in specific areas of the high school curriculum produced greater self-understanding among students in these areas.<sup>20</sup>

Related Research Conducted  
With College Freshmen

Holmes, working with college freshmen, compared the effectiveness of four techniques of test score reporting. Her subjects were 154 entering freshmen who had been administered a battery of tests including measures of intelligence, reading, personality, attitude, and interest. The sample was divided into four groups. The members of each group received a report of test results through one of the following methods: (1) the counselor described each test and pointed out the student's standing on a profile, answered questions, and made suggestions; (2) the counselor encouraged student participation, elicited self-estimates prior to reporting each score, and allowed the student to choose the order of the tests to be interpreted; the counselor maintained a non-evaluative attitude and focused on expressed affect regarding test results and self-ratings; (3) the counselor selected the order of test interpretation and indicated the student's results; student participation was elicited through reflection, clarification, and exploration of student

---

<sup>20</sup>Jacob Schulman, "A Comparison Between Ninth and Twelfth Grade Students on Self-estimates of Abilities and Objective Scores on the Differential Aptitude Tests," Dissertation Abstracts, Vol. XVI (1956), pp. 285-286.

feelings and attitudes toward the test results; (4) students were mailed a profile of their test results along with a brief summary of the meanings of the scores.

Among the conclusions drawn by the author were the following: (a) all four methods were effective; however, the students who were mailed their results considered them less valuable; (b) the method used by the counselor is less important than the counselor's personality; (c) students tended to rate themselves lower just prior to learning their actual test results than when they made their original estimates; and (d) the interpretive process improved student self-understanding in all four methods. There was no evidence that the use of the test profile in the interpretation process either increased or decreased the students' use and understanding of the information received.<sup>21</sup>

Wright compared group reporting of test results with individual reporting, both with a control group which received no report of results between pretest and post-test ratings. His sample contained 750 entering college freshmen, chosen at random, stratified by age and sex, and divided into three groups of 250 each. The group reporting technique included a general discussion of tests, their uses and limitations, and an explanation of the meaning of norms and percentiles. Following the discussion, each person

---

<sup>21</sup>June Elaine Holmes, "A Comparison of Four Techniques Used In Presenting Test Information to Freshmen Students," Dissertation Abstracts, Vol. XXI (1961), p. 3379.

received his own scores in quintiles on a slip of paper. This was followed by a period of time during which questions could be asked. The two counseled groups showed significant gains from the experience. Both groups had significantly greater gains than the control group, but there were no statistically significant differences in gain between individual and multiple counseling methods.<sup>22</sup>

Gustad and Tuma investigated the effects on student learning of three methods of test introduction and four methods of test interpretation with a sample of college freshmen. The methods of test introduction consisted of varying degrees of responsibility assumed by the counselor for introducing the idea for testing. The methods of test interpretation consisted of varying degrees of responsibility assumed by the counselor for pointing out discrepancies between students' self-estimates and test-estimates. The authors found no statistically significant differences in either the methods of test introduction or the methods of test interpretation on student learning. They concluded that student learning was positively related to initial accuracy of self-ratings but not to scholastic aptitude.<sup>23</sup>

---

<sup>22</sup>E. Wayne Wright, "A Comparison of Individual and Multiple Counseling for Test Interpretation Interviews," (Unpublished Doctoral Dissertation, University of California, Berkeley, 1957).

<sup>23</sup>John W. Gustad and A.H. Tuma, "The Effects of Different Methods of Test Introduction and Interpretation on Client Learning in Counseling," Journal of Counseling Psychology, Vol. IV, No. 4 (1957), pp. 313-317.

### Other Related Research

Froehlich sought to determine the effects of taking tests on self-ratings. He studied adult evening school students who rated themselves on seventeen abilities, interest, and personality traits before and after taking an extensive battery of tests. He found very few significant changes in self-ratings. It was concluded that the taking of tests, in and of itself, does not influence self-ratings and that these findings may be generalized to other studies where they act as a control over this one element in the total process. In conclusion, Froehlich states that whatever improvement there is in the accuracy of self-estimates would not be attributed to the experience of taking tests, but to other elements, in particular to receiving a report of the test results.<sup>24</sup>

### Summary of the Research

The following generalizations related to the problem of reporting test results to students, are summarized to serve as a base for the present research:

1. Test score reporting is probably closely related to other learning experiences and may be no more enduring than has been found to be the case with school subjects.
2. Test score reporting rarely increases student

---

<sup>24</sup>C.P. Froehlich, "Does Test Taking Change Self Ratings?", California Journal of Educational Research, Vol. V (1954), pp. 166-169; 175.

self-understanding below the high school level. Student maturity or readiness for this kind of self-information appears to be a significant variable.

3. Self-rating devices have been useful tools in assessing individual self-perceptions. Comparisons of student self-estimates and obtained test-estimates made prior to testing, following score reporting, and after specified time delays have provided useful measures of changes resulting from testing and score reporting to students.

4. The act of taking tests, in and of itself, does not influence the accuracy of self-ratings of test performance.

5. There is general agreement among the studies that self-estimates increase in accuracy following test interpretation and that, while there is some loss over time, there is significant improvement over self-estimates made prior to test interpretation.

6. There is now considerable evidence that reporting techniques have not been systematically related to the degree of increased accuracy of self-estimates.

## CHAPTER II

### THE PROBLEM AND EXPERIMENTAL DESIGN

#### Introduction

The studies reviewed in Chapter I provide only limited evidence of the benefits students derive from receiving a report of the results of standardized tests. More specifically, they provide very few research verified guidelines for effective test score reporting procedures to aid the counselor charged with the responsibility of increasing student self-understanding.

The present study which attempted to provide additional guidelines for counselors is characterized by five distinct features. First, it attempted to determine whether or not any one of three conditions of test score reporting was more effective with tenth grade students of different ability levels. Other studies have alluded to the influence of student ability on the accuracy of self-estimates of test performance; however, the results reported have been contradictory and inconclusive. Second, the use of a self-interpreting profile under controlled conditions was evaluated. Third, all test results were reported to the subjects as

percentile bands. Fourth, the concept of the relationship between ability and achievement was interpreted to the subjects. Other studies have avoided the interpretation of this concept because of norm differences and the difficulty of interpreting when a difference between ability and achievement is a significant one. The use of a correlated test battery, which included both an ability measure and tests of achievement in several subject areas and which reported obtained scores as percentile bands, facilitated the interpretation of the concept of over and under-achievement. Fifth, an attempt was made to evaluate the effect of modified instructions for the self-estimate criterion questionnaire on the basic findings of the study.

#### Statement of the Problem

This study was designed to investigate the influence of reporting conditions and the influence of student mental ability levels on the efficacy of test score reporting to tenth grade students. Specifically, does reporting in groups, reporting to individuals, or reporting test scores using only a "self-interpreting" folder produce significantly different results with students identified as having either low, middle, or high academic ability? Also, an attempt was made to determine whether the outcomes of the study would be significantly modified by changing the directions for marking the self-estimate questionnaire.

Operational Definitions

For purposes of this investigation significant terms to be used are defined in the following statements:

1. Criterion Instrument--questionnaire designed specifically to determine the degree of congruence between student self-estimates of test performance and his actual performance on Form 2A of the Sequential Tests of Educational Progress (STEP) and the School and College Ability Test (SCAT).
2. Test Score Reporting--the presentation of test results to students through an explanation of both the meaning of the scores in terms of percentile bands and defining the norm group with which the individual has been compared.
3. Recall Responses--recorded responses on the criterion instrument which resulted from a marking instruction that directed the individual to duplicate as nearly as possible his recorded test scores.
4. Acceptance Responses--recorded responses on the criterion instrument which resulted from a marking instruction that directed the individual to mark the questionnaire to reflect what he actually believed his achievement and abilities to be.



5. Ability Levels

- a. Low--subjects who obtained a total raw score of 50 or below on the SCAT.
- b. Middle--subjects who obtained total raw scores from 59 through 75 on the SCAT.
- c. High--subjects who obtained a total raw score of 83 or higher on the SCAT.

6. Reporting Conditions

- a. Group--a report of test results given to 25-30 students in a group setting. The presentation of these results included an oral and written explanation of their meaning.
- b. Individual--each subject in this reporting condition received a report of his test results in an individual interview.
- c. Student Report Folder (SRF)--a report of test results given to 25-30 students in a group setting. These results were presented on a profile which was part of a four page publisher's form. A special supplement was provided along with the profile. No oral explanation of the scores was given.
- d. Control--subjects received no report of their test results between the marking of the pre-test questionnaire and the follow-up questionnaire.

## 7. Abbreviations

- a. STEP--Sequential Tests of Educational Progress.
- b. SCAT (T)--School and College Ability Test--Total Raw Score.
- c. PT-Q--pretest questionnaire (criterion instrument marked prior to testing).
- d. IR-Q--immediate recall questionnaire (criterion instrument marked immediately following report of scores).
- e. IR(Gain)--immediate recall gain score.
- f. DR-Q--delayed recall questionnaire (criterion instrument marked one month following report of scores).
- g. DR(Gain)--delayed recall gain score.
- h. PT-DE--pretest, direction of error.
- i. IR-DE--immediate recall, direction of error.
- j. DR-DE--delayed recall, direction of error.
- k. INST-I--special marking instruction designed to elicit recall responses--control group only.
- l. INST-II--special marking instruction designed to elicit acceptance responses--control group only.
- m. RT-Q--retest questionnaire--criterion instrument marked one month after PT-Q--control group only.
- n. SRF--Student Report Folder--published by the Educational Testing Service for use with STEP/SCAT.

## Hypotheses

The following null hypotheses were formulated to determine whether or not reporting conditions have a significant

influence on the degree of congruence between self-estimates of test performance and actual scores obtained from tests:

Ho<sub>1</sub>: There are no statistically significant differences among means of gain scores of congruence for students who received their test results either in groups, individually, or through the use of a self-interpreting profile alone when the scores are obtained immediately following score reporting.

Ho<sub>2</sub>: There are no statistically significant differences among means of gain scores of congruence for students who received their test results either in groups, individually, or through the use of a self-interpreting profile alone when the scores are obtained one month after score reporting.

The following null hypotheses were formulated to determine whether or not student ability level has a significant influence on the degree of congruence between self-estimates of test performance and actual scores obtained from tests:

Ho<sub>3</sub>: There are no statistically significant differences among means of gain scores of congruence for students identified as having either low, middle, or high ability when the scores are obtained immediately following score reporting.

Ho<sub>4</sub>: There are no statistically significant differences among means of gain scores of congruence for students identified as having either low, middle, or high ability when the scores are obtained one month after score reporting.

The following null hypotheses were formulated to

determine whether or not the interaction between reporting conditions and mental ability levels has a significant influence on the means of gain scores of congruence:

Ho<sub>5</sub>: There is no statistically significant interaction between reporting conditions and student ability levels when data obtained immediately following the report of test results are analyzed.

Ho<sub>6</sub>: There is no statistically significant interaction between reporting conditions and student ability levels when data obtained one month after the report of test results are analyzed.

The following null hypothesis was formulated to determine whether or not altering the marking instructions for the self-estimate questionnaire would change the means of gain scores significantly:

Ho<sub>7</sub>: There is no statistically significant difference between the mean of gain scores obtained from an instruction designed to elicit recall responses and the mean of gain scores obtained from an instruction designed to elicit acceptance responses.

### Experimental Design

A brief outline of the test administration, sampling, and data collection procedures is presented to provide the reader with an overview of the experimental design:

1. First administration of the self-rating criterion instrument to the entire tenth grade population.

2. Administration of the ability and achievement test battery.

3. Selection of the sample based on SCAT total raw scores.

4. Reporting of test results to the subjects in the experiment.

5. Second marking of the criterion instrument (immediately following test score reporting session).

6. Retest administration of the criterion instrument to the control groups for reliability and control data.

7. Third marking of the criterion instrument by all but the control groups (one month following the test score reporting sessions).

8. Reporting of test results to the control group subjects.

9. Third administration of the criterion instrument to the control group subjects (immediately following the test reporting sessions).

10. Fourth administration of the criterion instrument to the control group subjects using an instruction designed to elicit recall responses (INST-I).

11. Fifth administration of the criterion instrument to the control group subjects using an instruction designed to elicit acceptance responses (INST-II).

A more detailed explanation of the preceding procedures will follow subsequently.

Figure 1 depicts graphically the twelve-cell two-part analysis of variance design for analyzing the influence of reporting conditions, ability levels, and the interaction between the two on the effectiveness of test score reporting with equal observations in each cell. Null hypotheses one through six were either accepted or rejected on the basis of the two-part analysis of variance described by Edwards.<sup>1</sup>

#### Instruments Used in the Study

The instruments used in this study consisted of (1) a self-rating scale, (2) a standardized achievement test battery, (3) a standardized academic ability test, (4) a test profile, and (5) a supplement to the test profile. The following paragraphs contain brief descriptions of the instruments.

#### The Self-Rating Scale

The self-rating scale used in this study was specifically designed for use with STEP and SCAT. Its purpose was to provide ratings of student self-estimates of (1) achievement, (2) ability, and (3) the relation between the two that could be directly compared with test-estimates of the same areas as measured by the STEP/SCAT test battery. Change in congruence between student self-estimates of test performance and actual performance on STEP/SCAT following test score

---

<sup>1</sup>Allen L. Edwards, Statistical Methods for the Behavioral Sciences (New York: Rinehart & Company, Inc., Second Printing, 1955), pp. 340-346.

Reporting Conditions	Ability Levels		
	Low	Middle	High
Group	Cell #1	Cell #2	Cell #3
	n=20	n=20	n=20
Individual	Cell #4	Cell #5	Cell #6
	n=20	n=20	n=20
SRF	Cell #7	Cell #8	Cell #9
	n=20	n=20	n=20
Control	Cell #10	Cell #11	Cell #12
	n=20	n=20	n=20

Fig. 1.--Twelve cell two-part analysis of variance design for analyzing the influence of reporting conditions, ability levels, and the interaction between the two on the effectiveness of test score reporting with equal observations in each cell.

reporting was the criterion used to determine the influence of student ability levels and the influence of reporting conditions on the effectiveness of test score reporting.

A copy of the criterion self-rating instrument appears in Appendix B, pages 103-106. The final form of the rating scale was prepared only after extensive tryouts, and several revisions were accomplished. The reading level of the directions and other content was checked with the Fog Index<sup>2</sup> and was found to be commensurate with the educational level of the sample used in this study. A pilot test-retest reliability study conducted with 65 subjects produced a Pearsonian correlation of .788 following a 20 day delay between administrations.

#### The Achievement Battery

The achievement battery selected for use in this study was the Sequential Tests of Educational Progress (STEP), Form 2A, published by the Educational Testing Service, Princeton, New Jersey. The main criteria for inclusion of this particular battery were:

(1) The STEP tests provide results sufficiently reliable and valid to serve as criteria of the achievement level attained by the subjects in each of the areas measured.

(2) The results of the STEP tests are reported as percentile bands which facilitated the interpretation of

---

<sup>2</sup>Edwin B. Flipppo, Principles of Personnel Management (New York: McGraw-Hill Book Company, Inc., 1961), p. 463.



significant difference between achievement and ability as measured by a correlated academic ability test.

The STEP battery included a seventy minute test in each of the following subject areas: Mathematics, Science, Social Studies, Reading, and Writing. The tests in the STEP series are intended primarily to be measures of developed abilities in five broad areas of education. Validity data are still largely unavailable; however, the publisher stated that, "Content validity is best insured by relying on well-qualified persons in constructing the tests, as was done for the STEP series."<sup>3</sup>

Reliabilities reported for the five STEP tests are the results of internal analyses based on single administrations of the tests. Reliabilities were estimated for four samples tested in the norms program. Analyses were done separately for grades five, eight, eleven, and thirteen. Only Form A was analyzed. Correlations between scores on alternate forms or between test-retest scores have not been obtained.

The following reliability coefficients are reported in the Technical Report for Form 2A, based on a sample of 110 students in grade eleven: Mathematics .84, Science .81, Social Studies .84, Reading .92, and Writing .85.<sup>4</sup>

---

<sup>3</sup>Sequential Tests of Education Progress-Technical Report (Princeton, New Jersey: Educational Testing Service, 1957), p. 9.

<sup>4</sup>Ibid., p. 10.

Intercorrelations obtained from a sample of 920 freshmen college students, using Form 1A of the five STEP tests ranged from .474 between Writing and Mathematics to .780 between Reading and Social Studies.<sup>5</sup>

The norming of the STEP tests was conducted simultaneously with the norming of the School and College Ability Test (SCAT). Fifty schools located throughout the nation provided a total norming sample of about 40,000 students in grades four through twelve.

#### The Academic Ability Test

The School and College Ability Test (SCAT), Form 2A, was selected for use in this study because the scores obtained from this test are directly comparable to the achievement tests in the STEP series.<sup>6</sup> SCAT was designed to estimate general ability to do school work. It provided a verbal ability (V) score, a quantitative ability (Q) score, and a total (T) score.

The publisher of SCAT, the Educational Testing Service, has conducted a number of validity studies which are reported in the 1958 STEP/SCAT Supplement.<sup>7</sup> Predictive validity was the central focus of these studies. For a more detailed account of the findings of these studies the reader is

---

<sup>5</sup>SCAT-STEP Supplement-1958 (Princeton, New Jersey: Educational Testing Service, 1958), p. 29.

<sup>6</sup>Op. cit., STEP-Technical Report, p. 19.

<sup>7</sup>Op. cit., SCAT-STEP-Supplement-1958, pp. 5-17.

referred to the publication mentioned above.

The reliabilities reported for SCAT are the results of internal analyses based on single administrations of the tests. They are, therefore, estimates of internal consistency. Correlations between alternate forms and test-retest correlations have not been reported.

Kuder-Richardson Formula 20 was used to estimate all of the reliabilities and standard errors of measurement. The reliability of SCAT, Form 2A, based on a sample of 2,292 students in grade eleven, is reported as follows: Verbal .92, Quantitative .90, and Total .95.<sup>8</sup>

#### The Test Profile

The four page Student Report Folder (SRF), published by the Educational Testing Service (ETS) as an aid to the interpretation of the results of the STEP/SCAT battery, was used to record test results for all subjects in all reporting conditions. A profile (SRF-page 2) was prepared for each student in advance of the reporting sessions. Results for the five STEP tests and the Verbal, Quantitative, and Total SCAT were presented as percentile bands. Each band was colored red. No raw scores, converted scores, or mid-percentile point scores were recorded on the profile. A sample copy of the SRF is included in Appendix B, pages 111-114.

---

<sup>8</sup>School and College Ability Test-Technical Report (Princeton, New Jersey: The Educational Testing Service, 1957), p. 11.

### The Supplement to the Test Profile

A special supplement to the SRF was prepared for use in this study. Examination of the SRF revealed that it did not provide a complete explanation of the meaning of the test results plotted on the profile. While it did explain adequately the purpose and content of the STEP/SCAT battery, it failed to provide the user with information concerning the meaning of several terms and concepts which are essential to a more complete understanding of the test results. The supplement was designed to alleviate this deficiency. The supplement was written at a level of readability suitable for the tenth grade subjects of this investigation. A sample copy of the supplement is included in Appendix B, pages 115-116.

### The Sample

The sample used in this investigation was drawn from the total population of 750 tenth grade students enrolled in a central Oklahoma urban secondary school. The school was chosen for the following reasons: (1) the enrollment was large enough to provide the number of subjects required by the design of the experiment; (2) the regular testing program of the school included the Sequential Tests of Educational Progress (STEP) and the School and College Ability Test (SCAT) for administration at the tenth grade level and; (3) the administrative and guidance staffs were amenable to cooperation with the research project.

Three groups of tenth-grade students, containing 80 students each, equally divided by sex, were identified as subjects for the study. One group, designated the low ability group, was drawn randomly from the 117 subjects who obtained a total raw score of 50 or below on the SCAT. This cutoff score was one standard deviation below the mean of the SCAT total raw score distribution for the 750 tenth grade students tested. A second group, designated the middle ability group, was drawn randomly from the 212 subjects whose total raw score on the SCAT was between 59 and 75 (plus and minus one-half standard deviation from the mean of the raw score distribution). The third group, designated the high ability group, contained 80 subjects drawn randomly from the 120 students who scored one standard deviation, or higher, above the mean of the local distribution of SCAT total raw scores. The cutoff raw score for the high ability group was 83.

A three by four, twelve cell design containing an equal number of subjects in each cell was obtained by assigning randomly by lot the 80 subjects within each of the three ability levels, to three test reporting treatment groups and a control group. Each cell contained twenty subjects--ten boys and ten girls. Following the assignment of the 240 subjects to the twelve cells, SCAT (T) raw score means and variances were computed for each cell of the matrix (see Table 25, Appendix A, page 93).

In order to determine whether or not the matching procedures had produced statistically homogeneous groups by ability levels, Bartlett's Test of Homogeneity of Variance<sup>9</sup> and the analysis of variance<sup>10</sup> were applied to the data. Table 1 summarizes the obtained chi-square and  $\underline{F}$  values for each of the three ability levels. Since none of the obtained chi-square values exceeded 7.815, it was concluded that the variances within each ability level were homogeneous. None of the obtained  $\underline{F}$  values exceeded 8.57 which led to the conclusion that the SCAT (T) raw score means within each of the three ability levels were not significantly different.

TABLE 1

SUMMARY OF BARTLETT'S TEST OF HOMOGENEITY OF VARIANCE  
CHI-SQUARE VALUES AND ANALYSIS OF VARIANCE  $\underline{F}$   
VALUES FOR GROUPS MATCHED FOR ABILITY  
LEVEL USING SCAT TOTAL RAW SCORES

Ability Levels	$\chi^2$ and $\underline{F}$ Values	
	$\chi^2$	$\underline{F}$
Low	4.388	2.745
Middle	2.721	2.073
High	3.692	1.960
	$\chi^2_{.05} = 7.815$	$F_{.05} = 8.57$

<sup>9</sup>Allen L. Edwards, Experimental Design in Psychological Research (New York: Rinehart & Company, Inc., 1960), pp. 125-126.

<sup>10</sup>Henry E. Garrett, Statistics in Psychology and Education, 5th ed. (New York: Longmans, Green and Co., Reprinted; 1960), pp. 280-281.

(Tables 23 and 24 in Appendix A, pages 91 and 92 present the data for this analysis in more detail.)

Since the analysis of the SCAT (T) raw score means within ability levels indicated homogeneity, the four means within each ability level were combined and the mean differences between levels were subjected to t test analysis. Combining four cells of 20 subjects each produced an "n" of 80 for each of the three ability levels. Table 2 summarizes the t test computations. All t values were significant at the .01 level. From this analysis it was concluded that the SCAT (T) raw score means were significantly different between all combinations of ability levels, i.e., the low group mean was significantly lower than the middle group mean, the low group mean was significantly lower than the high group mean, and the middle group mean was significantly lower than the high group mean.

TABLE 2

SIGNIFICANCE OF THE DIFFERENCE BETWEEN SCAT (T) RAW SCORE  
MEANS BY ABILITY LEVEL

Ability Level	Mean	Mean <sub>diff</sub>	SE <sub>diff</sub>	<u>t</u>
Low-Middle	41.5250 66.5375	25.0125	.9864	25.357*
Low-High	41.5250 89.7875	48.2625	1.0234	47.159*
Middle-High	66.5375 89.7875	23.2500	.7596	30.628*

\*Significant at .01 level

In summary, 80 subjects in each of three ability levels were randomly assigned to three treatment groups and a control group. The twelve groups, so identified, contained 20 subjects each--10 boys and 10 girls. Statistical tests for homogeneity of SCAT (T) means and variances were applied. No statistically significant differences between the means or between the variances were found within the ability levels. Mean differences between the three ability levels were found to be statistically significant at the .01 level.



## CHAPTER III

### COLLECTION OF THE DATA

#### Procedures for Administration of the Pretest Questionnaire and the Standardized Tests

The total population of students in grade ten were administered the criterion self-rating questionnaire. This was accomplished immediately prior to the administration of the SCAT and five tests of the STEP series. All students were tested simultaneously through the use of an intercommunications system. Directions for the administration of the questionnaire and SCAT/STEP were presented by an experienced test administrator. The students were located in rooms limited to 30 pupils each. A proctor who had received special orientation prior to the testing sessions was assigned to each room. The functions of the proctors were (1) to answer any questions which arose before each test was begun, (2) to check to see that answer sheets were properly prepared and marked, and (3) to see that each student worked on his own.

#### Scoring the Tests

After the administration of the pretest questionnaires and the standardized test battery, the completed

questionnaires and answer sheets were forwarded to the Department of Evaluation and Testing, Extension Division, University of Oklahoma, for scoring and processing. The SCATs were scored first, and a distribution of total raw scores was obtained (Appendix A, Table 22, page 90). The mean, median, standard deviation, and skewness were computed for the distribution of SCAT (T) raw scores. From these data the cutoff scores used in selecting the sample by ability levels were determined. A detailed explanation of the procedures used in selecting the sample was presented in Chapter II.

#### Recording Test-estimates and Scoring the Pretest Questionnaires for Congruence

After the subjects were identified, their STEP tests were removed from the total group of tests and scored. A data sheet especially designed for use in this investigation (Appendix B, page 110) was prepared for each subject. The column marked ACTUAL SCORES on the data sheet was completed by recording the results of STEP/SCAT according to the following procedures. The mid-percentile point score for each of the five STEP tests and the V, Q, and T scores of the SCAT were determined from the technical reports of the two tests. Each subject's scores were then recorded on his data sheet on the basis of their quintile location, i.e. mid-percentile scores from 0-19 were assigned a quintile value of 1, mid-percentiles from 20-39 a quintile value of 2, from 40-59 a

value of 3, from 60-79 a value of 4, and 80-99+ received a value of 5. The final five entries in the ACTUAL SCORES column were assigned values of 1, 2, or 3 on the following basis: A 1 was assigned if the percentile band for a specified ability as measured by SCAT was significantly higher than the percentile band for a specified STEP subtest, i.e. an indication of underachievement; a 2 was assigned if the ability and achievement percentile bands being compared overlapped, and a 3 was assigned if the SCAT measured ability was significantly lower than the STEP achievement area with which it was being compared, i.e. an indication of overachievement.

Following the recording of the test-estimates, the pretest questionnaires were scored for each of the subjects in the sample. Complete directions for scoring the self-rating scales are included in Appendix B, page 107. The score obtained from these procedures provided a pretest estimate of the degree of congruence between student self-estimates and obtained test-estimates for each of the subjects. A score indicating direction of self-estimate errors was also derived by following the scoring procedures outlined in Appendix B, page 108.

#### Preparations for Score Reporting

A Student Report Folder (SRF) was prepared for each subject in the investigation. The profile on page two of the SRF was completed according to the publisher's recommendations by marking off the percentile bands for each of the eight

scores obtained and coloring the bands red. A copy of the supplement to the SRF was inserted in the folder. These materials were then returned to the cooperating school in three groups. (The materials for the control group were retained at the scoring center.) One of the three groups of materials was designated to be used with the subjects which were to receive their results in group reporting sessions. One was designated to be used with the subjects who would receive their results on an individual basis, and the other group was to be used with students who would receive their results through the use of the reading materials only.

#### Score Reporting Procedures

At the time this research was planned it was determined that all reporting of test results, regardless of method, would be conducted by a single counselor. While it is realized that a given counselor may be more effective in the use of one method than another, it was believed that there would be fewer differentiating effects if only one counselor was used. When two or more counselors conduct the test reporting sessions, each applying his own preferred methods, any resulting differences in outcome may well be simply a reflection of differences in counselor competency. The decision to use just one counselor was made as an effort to control, as effectively as possible, the counselor variable. The decision to use the SRF and its special supplement under all reporting conditions was made as an attempt to control

the content of the reporting sessions.

The counselor regularly assigned counseling responsibility for the tenth grade conducted the reporting sessions. She was forty-six years old at the time the data were collected. She has a M. Ed. degree in counseling and guidance and holds a standard Oklahoma school counselor certificate. She has three years business experience, seven years classroom teaching experience, and has been a counselor for the past two years.

The 240 subjects were randomly assigned to three treatment groups and a control group for score reporting. One-fourth of the sample received their STEP/SCAT test results in groups of 25-30. Each student was provided with his SRF (with supplement). The counselor then presented a generalized interpretation of the purpose and content of the tests, the norm group with which the subjects were being compared, and the meaning of the results. The SRF and the supplement were used as a guide for this presentation. (A complete typescript of a group reporting session is included in Appendix C, pages 118-131.) Following the general explanation, individual questions were answered and the folders were re-collected.

A second one-fourth of the sample received their test results on an individual basis. Each subject was called to the counselor's office where he was provided with his SRF (with supplement). The counselor and the student then went

over the results together. All questions asked by the student regarding the results were answered. (A complete typescript of an individual session is included in Appendix C, pages 132-151.)

Another one-fourth of the sample were called together in groups of 25-30 where each subject had his SRF (with supplement) given to him with instructions to read the materials carefully in order to understand the meaning of his results. The subjects in this reporting condition were instructed to read the materials in the following order: First, page one of the SRF; second, the supplement; third, page four of the SRF; and finally, the inside two pages of the SRF which contained the profile of results with a description of the content of each of the STEP/SCAT tests. The subjects were given 35 minutes to examine their folders before they were re-collected. The counselor remained in the room during the entire time period and was available to answer individual questions.

The remaining one-fourth of the sample were used for a control and received no report of their test results. The subjects in this category were called together thirty days following the administration of the pretest questionnaire (PT-Q) and requested to mark the same instrument a second time (RT-Q).

Marking the Self-estimate  
Criterion Instrument

Immediately following the re-collection of the SRFs in each reporting session all subjects were administered the self-estimate criterion instrument developed for use in this study. Results from the scoring of this form provided an estimate of the degree of congruence between student self-estimates of test performance and actual test scores immediately following test score reporting. The difference between the pretest questionnaire (PT-Q) score and the immediate recall questionnaire (IR-Q) score was regarded as an estimate of the gain (IR-Gain) in self-understanding derived from the test reporting experience.

Thirty days (plus or minus five days) following the report of test results the subjects were reassembled for a third marking of the criterion instrument (DR-Q). Results from the scoring of this third marking were compared with the scores from the pretest questionnaire to determine an estimate (DR-Gain) of the retention of self-information acquired in the test reporting sessions.

Special Marking Instructions Used  
with the Control Group

As described previously, the three groups of 20 subjects each which were designated the control group were administered the pretest questionnaire (PT-Q), the STEP/SCAT test battery, and a retest questionnaire (RT-Q) thirty days following the administration of the PT-Q. Control and

reliability data were obtained by these procedures.

In order to determine whether the outcomes of the study would be significantly altered by modifying the marking instructions for the self-estimate criterion questionnaire, the procedures that follow were implemented. Two weeks after the second marking of the criterion questionnaire the subjects of the control group were assembled in two groups of 30 students each for a group test result reporting session. The procedures outlined above for the regular group reporting sessions were followed in detail. (Appendix C, pages 118-131, for a complete typescript of a group test result reporting session.) As soon as the explanation of results was completed the SRFs were re-collected, and a copy of the self-estimate questionnaire was provided to each subject. The students were then instructed to mark the instrument following the standard instructions (IR-Q) printed on the cover page of the questionnaire. When this was completed, the instruments were collected, and a second copy of the questionnaire was distributed. This time the subjects were requested to mark the instrument according to the following instruction (INST-I):

INSTRUCTIONS--for filling out questionnaires marked  
"I".

Please fill out this questionnaire to duplicate as nearly as possible your recorded test scores as you remember them from your Student Report Folder.

If this is the way you marked the questionnaire you just completed, then mark this one the same way.

(If not, refer to sentence one above.)



After completion of the questionnaire according to the above instruction, the instruments were collected, and a third copy was distributed. The subjects were then requested to mark this final questionnaire according to the following instruction (INST-II):

INSTRUCTIONS--for filling out questionnaires marked "II".

Please mark this questionnaire to express what you feel your actual achievement and abilities to be.

If you feel that your scores, as recorded in the Student Report Folder, were an accurate assessment of your achievement and abilities, then you should mark it just like you marked the previous questionnaire.

If you disagree with any or all of the scores reported in the Student Report Folder, mark the questionnaire so as to reflect what you feel your actual achievement and abilities are.

INST-I was designed to elicit recall responses (the accuracy with which the subject can recall his test results following the reporting session). INST-II was designed to elicit acceptance responses (the degree to which the subject accepts or rejects the results of his tests). Data obtained from the scoring of the three sets of questionnaires which were marked according to the procedures outlined above were analyzed to determine the influence of criterion instrument marking instructions on the relative congruence between student self-estimates of test performance and obtained test scores.

## CHAPTER IV

### STATISTICAL ANALYSES

#### Introduction

The primary statistical technique used to analyze the data obtained through the procedures outlined in Chapter III was the two-part analysis of variance described by Edwards.<sup>1</sup> This method of analysis was useful and appropriate to determine simultaneously the interaction of two or more variables, in this case, reporting conditions by ability levels. The application of analysis of variance procedures was predicated on two major underlying assumptions. These assumptions were: (1) the treatment groups used in the study were selected at random from the same population, and (2) the variances between the groups of data obtained were statistically homogeneous.<sup>2</sup> The assumption of randomness was satisfied by the sample selection procedures described in Chapter II and

---

<sup>1</sup>Edwards, Statistical Methods for the Behavioral Sciences, pp. 340-349.

<sup>2</sup>E.F. Lindquist, Design and Analysis of Experiments in Psychology and Education (Boston: Houghton-Mifflin, 1953), p. 73.

homogeneity of variance was checked by applying Bartlett's<sup>3</sup> test to the data. Other statistical methods used included t tests and correlation.

#### Analysis of Pretest Questionnaire (PT-Q) Data

After the subjects were identified and the standardized tests were administered and scored, the pretest criterion questionnaires (PT-Q) were scored for congruence between self-estimate scores and obtained test scores. Means and variances by ability levels and conditions of reporting were computed from the PT-Q data (Table 26, Appendix A, page 94).

In order to determine if statistically significant differences existed among means and variances of the PT-Q scores prior to the reporting of test results to the subjects, the data were submitted to an analysis of variance. Although Bartlett's test produced a chi-square value of 21.762, significant at the .05 level, which indicated that the variances were heterogeneous, the analysis of variance was completed. Lindquist suggested that:

. . ., the heterogeneity must be quite extreme to be of any serious consequence. While statistical tests of heterogeneity of variance are available . . ., there will be relatively few situations in which any such test is required. In general, unless the heterogeneity of either form or variance is so extreme as to be readily apparent upon inspection of the data, the effect upon the F-distribution will probably be negligible. In general, when the heterogeneity in form or variance is "marked" but not "extreme", allowance may be made for this factor by setting a

---

<sup>3</sup>Edwards, Experimental Design, pp. 125-126.

higher "apparent" level of significance for the tests of treatment effects than would otherwise be employed.<sup>4</sup>

Table 3 presents a summary of the two-part analysis of variance applied to the PT-Q data. Since none of the  $F$  values obtained were significant at the .05 level, the inability of the data to meet the assumption of homogeneity of variance was not considered pertinent in this case. From these analyses it was concluded that ability level did not have a statistically significant effect on the congruence between student self-estimates of test performance and obtained test scores prior to test score reporting.

TABLE 3  
SUMMARY ANALYSIS OF VARIANCE FOR THE  
PRETEST QUESTIONNAIRE DATA

Source of Variation	$\Sigma$ of Squares	df	Mean Square	F
Reporting Conditions	66.267	3	22.089	3.010
Ability Levels	300.400	2	150.200	2.259
Interaction	381.933	6	63.656	1.045
Between Groups	748.600	11	68.054	1.023
Within Groups	15161.000	228	66.496	
Total	15909.600	239		

<sup>4</sup>Lindquist, op. cit., p. 86.

### Analysis of Control and Reliability Data

The three groups of 20 subjects each which were designated the control group (total  $n=60$ ) were administered the PT-Q, the STEP/SCAT battery, and a retest questionnaire (RT-Q) thirty days following the administration of the PT-Q. Control and test-retest reliability data were obtained by analyzing the scores from these two administrations. The two sets of criterion questionnaire scores were correlated and produced a Pearsonian correlation of .696. The mean from the RT-Q data was .667 points higher than the PT-Q mean. This mean difference was tested for statistical significance and produced a  $t$  value of .930 which was not significant at the .05 level. From the results of these analyses it was concluded: (1) that the reliability of the thirteen item criterion questionnaire was acceptable and (2) that students do not improve their degree of congruence between self-estimates of test performance and actual test performance without a report of their results.

### Analysis of the Immediate Recall (IR-Q) Data

The congruence scores obtained from the PT-Qs formed the basis from which changes in congruence between self-estimate scores and obtained test scores were calculated following test score reporting. The second marking of the criterion instrument was accomplished immediately following the reporting sessions, and the data obtained were designated

immediate recall (IR-Q) data. The IR-Q score for each subject was compared with his PT-Q score, and a gain (IR-gain) score was obtained by subtraction. The following analyses were based on the means of gain scores for the nine experimental treatment groups (excluding the control groups).

Table 27, Appendix A, page 95, presents the means and variances of gain scores obtained from the IR-Q data.

Figure 2 is a graphic presentation of the IR-Q means of gain scores plotted by reporting conditions and ability levels. Inspection of the graph suggests that individual reporting may be superior to the other two reporting conditions for the low and middle ability groups but not for the high ability group. Ability level appears to be an important factor with respect to the amount of gain in congruence obtained regardless of reporting condition, i.e. there was no "overlapping" between ability levels.

In order to determine whether or not the differences among IR means of gain scores were statistically significant, Bartlett's test and the analysis of variance procedures were applied. Bartlett's test produced a chi-square value of 8.792 which was not statistically significant at the .05 level. Table 4 presents a summary analysis of variance for the IR data.

The between groups F value of 4.889 with 8 and 171 degrees of freedom was significant at the .01 level. The F values obtained for reporting methods and interaction

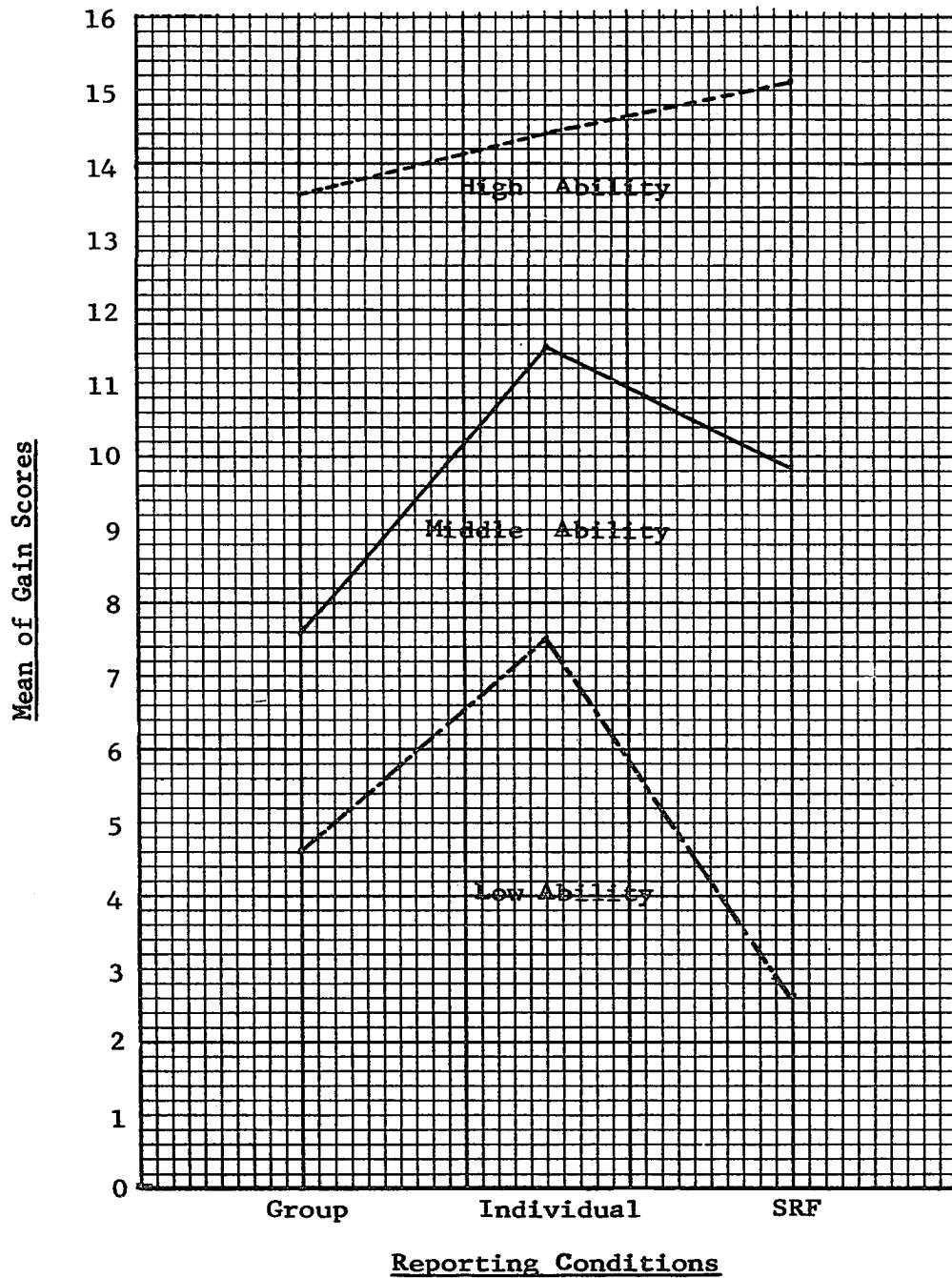


Fig. 2.--Immediate recall means of gain scores graphed by reporting conditions and ability levels.

TABLE 4

## SUMMARY ANALYSIS OF VARIANCE FOR THE IMMEDIATE RECALL DATA

Source of Variation	$\Sigma$ of Squares	df	Mean Square	F
Reporting Conditions	216.310	2	108.155	1.357
Ability Levels	2688.710	2	1344.355	16.879*
Interaction	210.224	4	52.556	1.515
Between Groups	3115.244	8	389.406	4.889*
Within Groups	13619.400	171	79.646	
Total	16734.644	179		

\*Significant at .01 level

indicated non-significance. The F value obtained for ability levels was significant at the .01 level. Since the F value for reporting conditions was not significant, the data for reporting conditions were pooled by each ability level for further analysis. This increased the size of "n" from three groups of 20 each to one group of 60 subjects. Table 28, Appendix A, page 96, presents the IR-gain data consolidated by each ability level. Table 5 summarizes the results of applying t test analysis to determine the significance of the differences between the means of gain scores for the three ability levels. Differences in means of gain were statistically significant at the .01 level for all combinations of ability level. The middle ability level gained significantly more from the test reporting sessions than did the low



ability groups and the high ability groups gained significantly more from the experience than did the middle ability groups.

TABLE 5  
SIGNIFICANCE OF THE DIFFERENCES BETWEEN MEANS OF GAIN  
FOR IMMEDIATE RECALL DATA BY ABILITY LEVEL

Ability Level	Means of Gain		Mean of Gain <sub>diff</sub>	SE <sub>diff</sub>	t
Low-Middle	4.900	9.700	4.800	1.595	3.009*
Low-High	4.900	14.367	9.467	1.734	5.460*
Middle-High	9.700	14.367	4.667	1.595	2.926*

\*Significant at .01 level

On the basis of these analyses null hypothesis one (there are no statistically significant differences among means of gain scores of congruence for students who received their test results either in groups, individually, or through the use of self-interpreting profile alone when the scores are obtained immediately following score reporting) was accepted. Null hypothesis five (there is no statistically significant interaction between reporting conditions and student ability levels when data obtained immediately following the report of test results are analyzed) was also accepted. Null hypothesis three (there are no statistically significant differences among means of gain scores of congruence for students identified as having either low, middle, or high ability when

the scores are obtained immediately following score reporting) was rejected.

Based on data obtained immediately following test score reporting, it was concluded that the change in congruence between student self-estimates of test performance and obtained test scores is not significantly influenced by either the reporting condition or the interaction between reporting conditions and ability levels; however, the change in congruence is significantly influenced by student ability level.

#### Analysis of the Delayed Recall (DR-Q) Data

The same procedures that were applied to the IR-Q data were used in the analysis of the DR-Q data. DR(gain) scores were computed by subtraction for each subject by comparing his PT-Q congruence score with his DR-Q congruence score which was obtained 30 days (plus or minus five days) after receiving a report of his test results. Table 30, Appendix A, page 97, presents the delayed recall means and variances of gain scores for the nine treatment groups. Figure 3 presents the DR means of gain scores graphed by reporting conditions and ability levels. Inspection of the figure suggests that individual reporting is retained to a higher degree than either of the other two conditions. The gain from individual reporting by the low ability group actually exceeds slightly the gain made by the middle ability group from the group reporting condition. The low ability

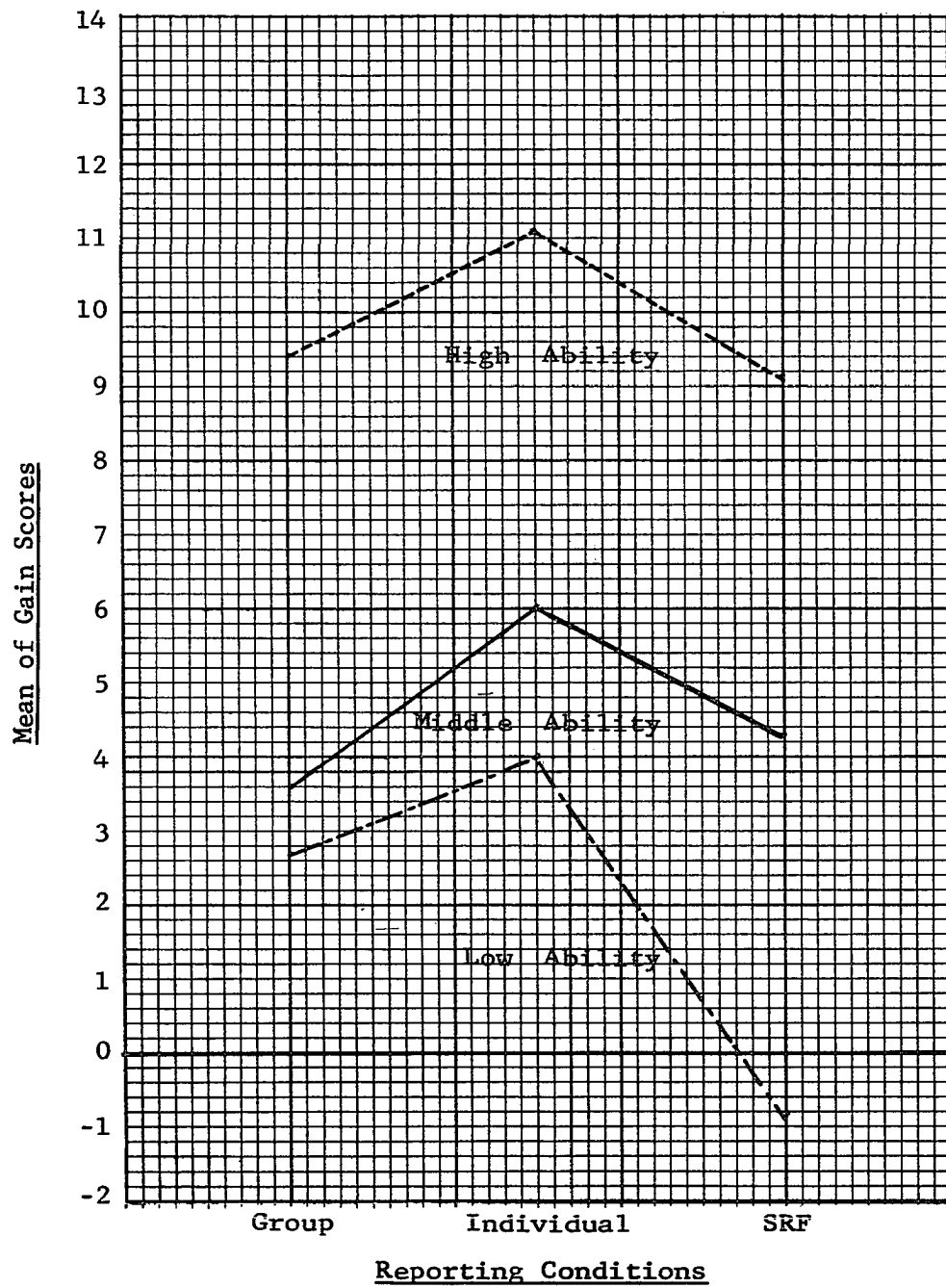


Fig.3.--Delayed recall means of gain scores graphed by reporting conditions and ability levels.

group's mean of gain score receded to a point below zero gain for the SRF reporting condition which indicates that presenting test results to low ability students using only reading materials is probably fruitless.

In order to determine whether or not the differences among DR means of gain scores were statistically significant, Bartlett's test and the analysis of variance procedures were applied. Bartlett's test produced a statistically non-significant chi-square value of 12.933. A summary of the analysis of variance procedures applied to the DR data is presented in table 6.

TABLE 6  
SUMMARY ANALYSIS OF VARIANCE FOR  
THE DELAYED RECALL DATA

Source of Variation	$\Sigma$ of Squares	df	Mean Square	F
Reporting Conditions	251.911	2	125.956	2.048
Ability Levels	1952.311	2	976.556	15.878*
Interaction	113.289	4	28.322	2.172
Between Groups	2317.511	8	289.689	4.710*
Within Groups	10517.400	171	61.505	
Total	12834.911	179		

\*Significant at .01 level

The between groups F value of 4.710 was statistically significant at the .01 level. The reporting conditions and interaction F values failed to reach statistical significance at the predesignated (.05) level. Ability levels reached significance at the .01 level. Since the F value for the reporting conditions was not significant, the data for reporting conditions were pooled by each ability level for further analysis. Table 29, Appendix A, page 96, presents the DR-gain data consolidated by each ability level. Table 7 summarizes the results of applying t test analysis to determine the significance of the differences between the means of gain scores for the three ability levels. Differences between means of gain were statistically significant for all combinations of ability level. The low ability subjects gained less than the middle ability subjects at the .05 level of significance. The low ability subjects gained less than the high ability subjects at the .01 level of significance and the middle ability subjects gained significantly less than the high ability subjects at the .01 level.

On the basis of these analyses null hypothesis two (there are no statistically significant differences among means of gain scores of congruence for students who received their test results either in groups, individually, or through the use of a self-interpreting profile alone when the scores are obtained one month after score reporting) was accepted. Hypothesis six (there is no statistically significant

interaction between reporting conditions and student ability levels when data obtained one month after the report of test results are analyzed) was also accepted. Hypothesis four (there are no statistically significant differences among means of gain scores of congruence for students identified as having either low, middle, or high ability when the scores are obtained one month after score reporting) was rejected.

TABLE 7

SIGNIFICANCE OF THE DIFFERENCES BETWEEN MEANS OF GAIN SCORES FOR DELAYED RECALL DATA BY ABILITY LEVELS

Ability Level	Means of Gain		Mean of Gain <sub>diff</sub>	SE <sub>diff</sub>	t
Low-Middle	1,933	4.633	2.700	1.309	2.062**
Low-High	1.933	9.867	7.934	1.450	5.471*
Middle-High	4.633	9.867	5.234	1.554	3.368*

\*Significant at .01 level

\*\*Significant at .05 level

Based on data obtained one month following test score reporting to students, it was concluded that the change in congruence between student self-estimates of test performance and obtained test scores is not significantly influenced by either the reporting condition or the interaction between reporting condition and ability level; however, the change in congruence is significantly influenced by student ability level.

Figure 4 presents the PT, IR, and DR means of gain

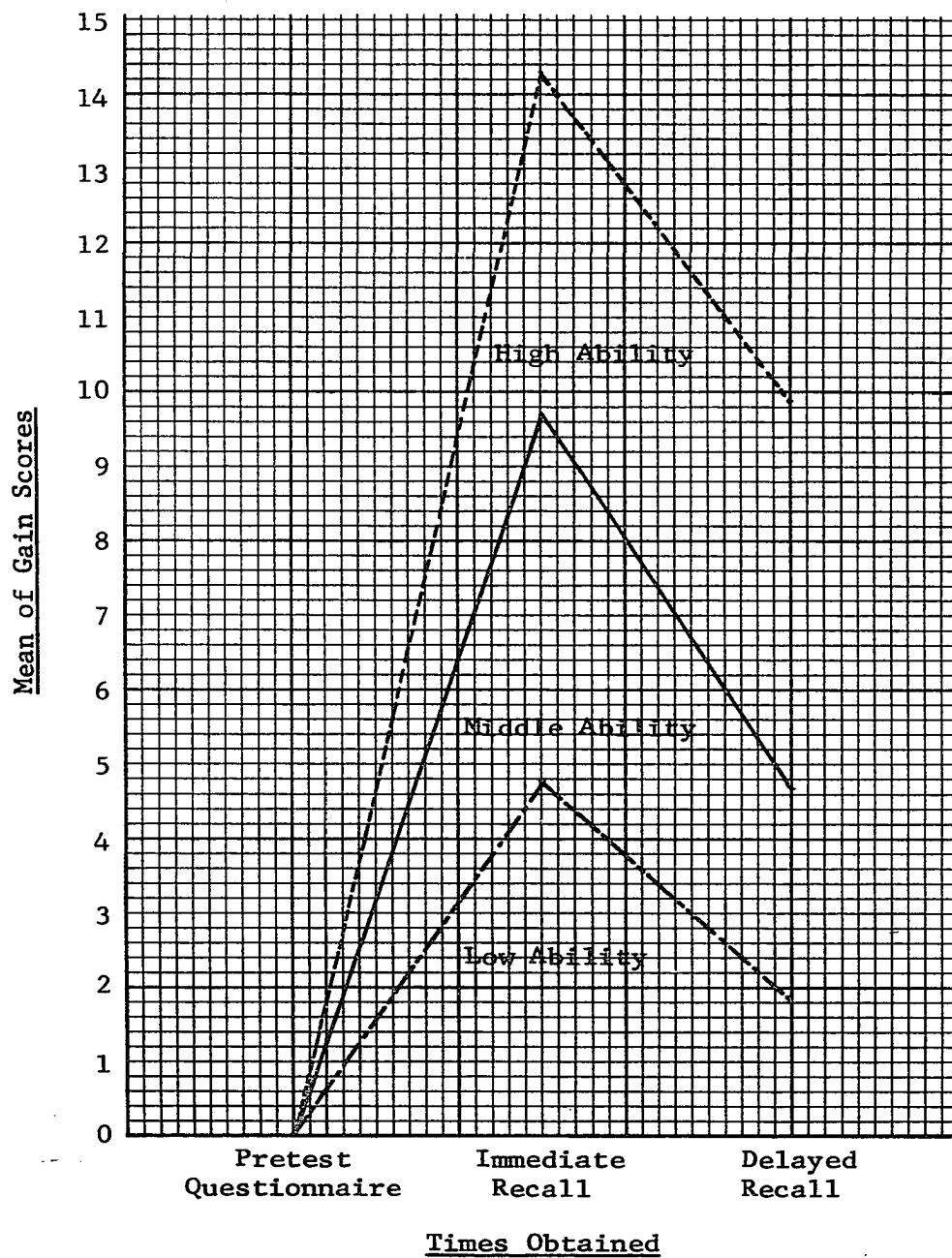


Fig.4.--Means of gain scores graphed by ability levels and times obtained.

scores consolidated by ability levels and graphed according to the time the data were obtained. The figure clearly depicts the influence of student ability on the degree of gain in congruence between self-estimates and test-estimates and the amount of the gain that is retained after a delay of one month.

Analysis of Data Obtained from Modifying  
the Marking Instructions for the  
Criterion Questionnaire

Two weeks after control and test-retest reliability data were obtained on the subjects in the control groups, these 60 students were assembled in two groups of 30 each and were presented a group report of their test results. Procedures developed for the regular group reporting sessions were employed. As soon as the reporting procedures were completed, the SRFs were collected, and the subjects marked the IR-Q according to the instructions printed on the questionnaire. When this was completed, the instruments were collected, and a second copy of the questionnaire was distributed. The subjects were requested to mark this instrument following a special marking instruction designed to elicit recall responses (INST-I). After completion and collection of the second questionnaire, a third copy of the instrument was distributed. This copy was marked according to a special instruction designed to elicit acceptance responses (INST-II).

The data obtained from the scoring for congruence of these three sets of questionnaires were analyzed to determine



whether or not changing the marking instructions for the criterion instrument would significantly alter the outcomes of the investigation. Correlation and t tests were used in the analysis.

Table 31, Appendix A, page 98, presents means, variances, and standard deviations of gain scores by ability levels and marking instructions. Figure 5 is a graphic presentation of the means of gain by ability levels and marking instructions. Inspection of Figure 5 suggests that INST-I (recall) elicits a higher degree of congruence between student self-estimates and obtained test-estimates than either IR-Q or INST-II. IR-Q and INST-II appear to produce about the same degree of congruence (mean of gain) within each ability level. The middle ability group achieved approximately the same mean gain from INST-I as did the high ability group.

Pearsonian product-moment correlations computed between the three sets of congruence scores for the 60 subjects of the control group are presented in Table 8.

TABLE 8

CORRELATIONS BETWEEN THREE DIFFERENT MARKING  
INSTRUCTIONS-CONTROL GROUP ONLY

Marking Instructions	Correlation
IR-Q--INST-I	.733*
IR-Q--INST-II	.772*
INST-I--INST-II	.725*

\*Significant at .01 level

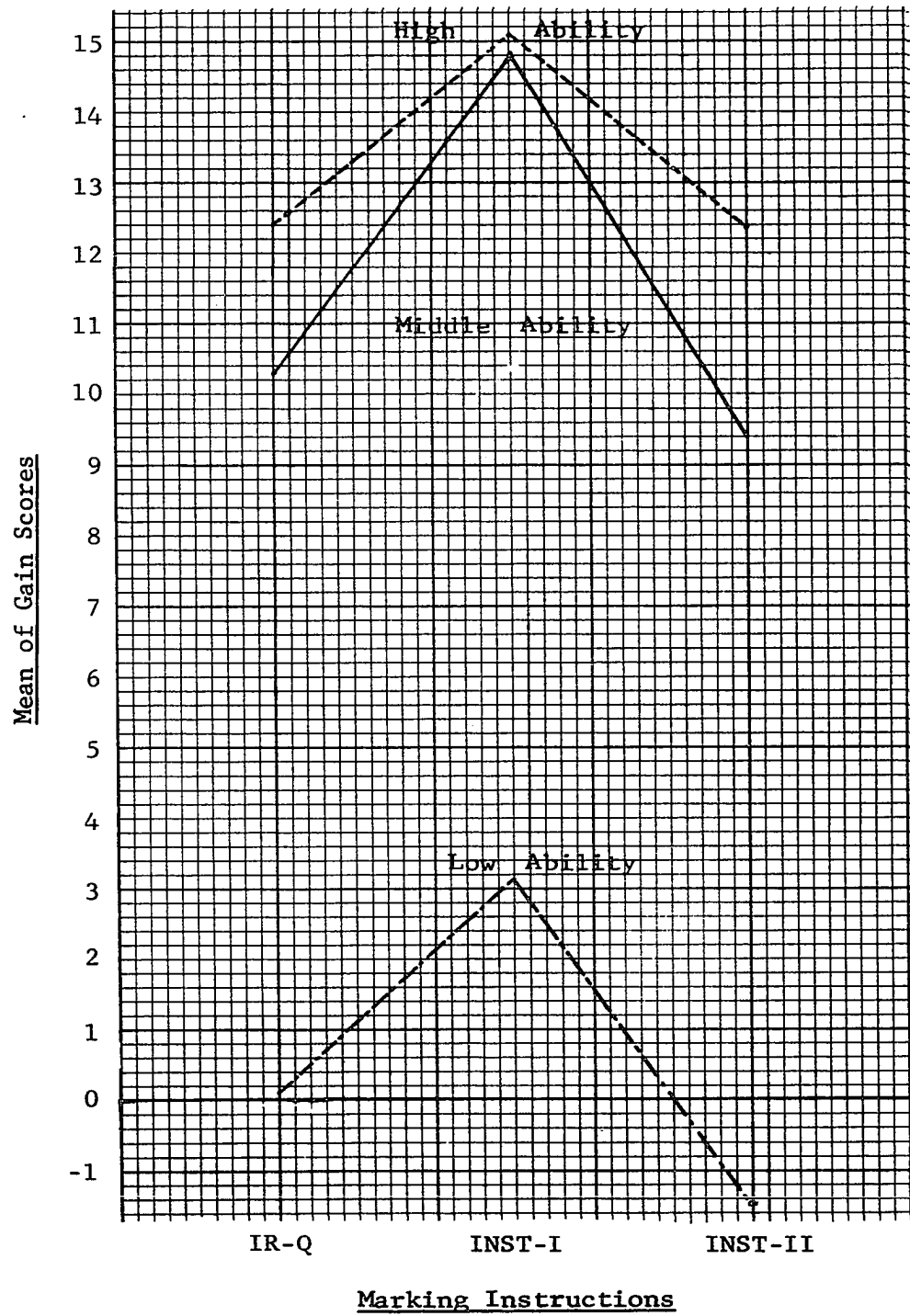


Fig.5.--Means of gain scores graphed by ability levels and marking instructions.

All of the obtained correlations varied significantly from zero at the .01 level of significance. Standard errors of the difference between each of the correlated pairs were obtained by using Fisher's<sup>5</sup>  $z$  function. Critical ratios were computed for each of the pairs, however, none were found to be significant at the .05 level.

Significance of the differences between means of gain scores obtained from the three marking instructions were analyzed through the use of  $t$  tests. A summary of the results by ability level and for the total group is included in Appendix A, Table 33, page 100. Table 9 presents these data for the total group (n=60).

TABLE 9

SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEANS OF GAIN  
SCORES WHEN THREE DIFFERENT MARKING INSTRUCTIONS  
WERE USED--CONTROL GROUP ONLY

Instruction	Means of Gain		Mean of Gain <sub>diff</sub>	SE <sub>diff</sub>	$t$
IR-Q vs INST-I	7.667	11.033	3.366	.944	3.566*
IR-Q vs INST-II	7.667	6.767	.900	.868	1.036
INST-I vs INST-II	11.033	6.767	4.266	.979	4.358*

\*Significant at .01 level

<sup>5</sup>Garrett, op. cit., pp. 199-202.

A statistically significant difference at the .01 level was obtained between the mean of gain for the IR-Q instruction and the mean of gain for the instruction designed to elicit recall (INST-I). The same statistical difference was obtained between INST-I (recall) and INST-II (acceptance). No statistically significant difference was found between the IR-Q instruction mean of gain and the INST-II mean of gain.

Null hypothesis seven (there is no statistically significant difference between the mean of gain scores obtained from an instruction designed to elicit recall responses and the mean of gain scores obtained from an instruction designed to elicit acceptance responses) was rejected on the basis of this analysis.

Further examination of Figure 5 suggests, however, that the basic outcomes of the study, i.e. the influence of reporting method and student ability level on the effectiveness of score reporting, would remain essentially unchanged regardless of which instruction was used. As the instruction was altered the mean of gain for each of the three ability levels changed in the same direction, i.e. the recall instruction (INST-I) produced an increase in mean of gain scores for all ability levels. Except for the near coincidence of the mean of gain scores for the middle and high groups on INST-I, the three ability levels maintained essentially the same relationship with each other that was reported from the IR and DR data.

### Further Analyses and Conclusions

Each time a self-estimate questionnaire was scored a direction of error score was computed. (Scoring instructions for direction of error are included in Appendix B, page 108.) Tables 10, through 21, pages 78-89, and table 34, page 101 in Appendix A present the data obtained from scoring for direction of error. Inspection of the tables confirms a trend reported by Froehlich and Moser<sup>6</sup> that students who score low on standardized tests remember their scores as being higher than they actually were while high scoring students remember their scores as being lower than was the actual case.

Table 32, Appendix A, page 99 is a presentation of the intercorrelations between the five different markings of the criterion instrument by the control group. The significant correlations between the pretest questionnaire (PT-Q) and subsequent markings (IR-Q and INST-II) confirm the findings of Gustad and Tuma<sup>7</sup> that student learning (from test reports) is positively related to initial accuracy of self-ratings.

---

<sup>6</sup>Froehlich and Moser, loc. cit.

<sup>7</sup>Gustad and Tuma, loc. cit.

## CHAPTER V

### SUMMARY, CONCLUSIONS, AND IMPLICATIONS

#### Summary

This study was designed to investigate the influence of reporting condition and mental ability level on the effectiveness of test score reporting to tenth grade students. Specifically, it was desired to determine whether or not any one of three conditions of test score reporting: reporting in groups, reporting to individuals, or reporting results using only a "self-interpreting" folder would be significantly more effective with students identified as having either low, middle, or high academic ability. Whether or not the outcomes of the study would be significantly modified by changing the directions for marking the self-estimate questionnaire was also investigated.

The subjects who participated in the study were 240 students drawn from the total population of 750 tenth-graders enrolled in a central Oklahoma urban secondary school. Eighty subjects in each of three ability levels as identified by the School and College Ability Test total raw score were randomly assigned to three treatment groups and a control group for

test result reporting. The twelve groups contained twenty subjects each--ten boys and ten girls.

A criterion self-rating questionnaire was administered before testing, immediately following test result reporting, and one month later to determine the relative degree of congruence between each student's self-estimates of achievement and ability, the relationship between the two, and his actual performance on Form 2A of the Sequential Tests of Educational Progress and the School and College Ability Test. The data obtained from the three markings of the criterion instrument by the treatment groups were analyzed through the use of two-part analysis of variance and t test procedures. The control group subjects were used later in the study to determine what effect the changing of the marking instructions for the criterion instrument would have on the findings of the investigation.

### Conclusions

Based on the results of the analyses of the data and with consideration for the limitations imposed by the design of the experiment, the use of a single counselor, and the instruments used for the collection of data, the following conclusions are suggested:

1. For tenth-grade students without previous exposure to the results of standardized test information, ability level does not influence the degree of congruence between student self-estimates of test performance and obtained test scores

prior to their receiving a report of their test results.

2. Tenth grade students do not significantly increase their degree of congruence between self-estimates and obtained test-estimates without a report of their test results.

3. The change in the degree of congruence between student self-estimates and obtained test-estimates immediately following the receiving of a report of test results is not significantly affected by either the reporting condition or the interaction between reporting conditions and ability levels.

4. Student ability level has a statistically significant effect on the change in degree of congruence between student self-estimates and obtained test-estimates when the change in congruence is measured immediately following the report of test results.

5. The change in the degree of congruence between student self-estimates and obtained test-estimates one month following the receiving of a report of test results is not significantly affected by either the reporting condition or the interaction between reporting conditions and ability levels.

6. Student ability level has a statistically significant effect on the change in degree of congruence between student self-estimates and obtained test-estimates when the change in congruence is measured one month following the report of test results.



7. Means of gain scores of congruence between student self-estimates and obtained test-estimates may be significantly changed by altering the marking instructions for the self-rating questionnaire.

8. The basic conclusions regarding the influence of reporting conditions and ability levels remains essentially unchanged regardless of the instruction used for marking the criterion instrument.

9. Subjects identified as having low ability tend to rate themselves as being higher than their obtained test-estimates suggest, while high ability subjects tend to under-rate themselves as compared to obtained test-estimates. These discrepancies in self-knowledge are reduced in magnitude following the receipt of test result information.

10. Student learning from test information is positively related to initial accuracy of self-ratings.

11. Student short-term learning of self-information through test result reporting is greater than the long-term effects, i.e. a substantial portion of the congruence between student self-estimates and obtained test-estimates that is gained immediately following the reporting of test results is not retained when remeasured on month later. The amount retained, however, is significantly greater than zero. This conclusion suggests that this kind of learning is probably no more enduring than has been found to be the case in retention studies of school subject material.

Implications for Counselors

An evaluation of the results of the present study suggests the following implications for counselors who are charged with the responsibility for increasing student self-understanding through the use of the results of standardized tests:

1. It was concluded in the present study that reporting conditions, as described, had no significant effect on the efficacy of test result reporting. This suggests that counselors may find that considerable time may be saved by reporting test results to students in well planned group sessions. Interpretive aids written at the reading level of the students involved should be provided, individual questions should be answered, and an invitation to visit the counselor on an individual basis should be extended to those who desire to do so.

2. There is substantial evidence from this and other studies that students in grades below nine or ten do not gain significantly from reports of their standardized test results. This phenomenon may be a manifestation of lack of maturity, lack of readiness for self-information, lack of ability to understand the meaning and/or significance of test scores, or some degree of interaction between all three. From this generalization, it would follow that little will be gained from reporting results of standardized tests to students below these grade levels, especially to those students whose ability level has been determined to be below the average of the

national distribution.

3. The data obtained for use in this study indicated that some individuals, regardless of ability level, did not change in degree of congruence between self-estimates and obtained test-estimates following test interpretation. Often the measured divergence was extreme. Counselors who wish to identify these individuals may do so by following up test reporting sessions with a self-rating questionnaire similar to the one used in this study and scoring it for self-estimate--test-estimate congruence.

#### Recommendations for Further Research

On the basis of the experiences gained and problems encountered while conducting the present research, the following recommendations are suggested for future research in the area of test result reporting to secondary school students:

1. Future research should expand the present effort to differentiate between recall and acceptance of test results.
2. Future research should be designed to identify and interpret the influence of student readiness for self-information.
3. Future research should be designed to determine the influence of ability level versus general maturity level on the acceptance of standardized test information, i.e. the difference in gain from test information between high ability ninth-graders versus low-ability twelfth-graders.

## BIBLIOGRAPHY

### Books

- Anastasi, Anne. Psychological Testing. New York: The MacMillan Co., 1956.
- Chauncey, Henry, and Dobbin, John E. Testing--Its Place in Education Today. New York: Harper and Row, 1963.
- Cronbach, Lee J. Essentials of Psychological Testing. 2d ed. New York: Harper & Brothers, 1960.
- Edwards, Allen L. Experimental Design in Psychological Research. New York: Rinehart & Company, Inc., 1960.
- \_\_\_\_\_. Statistical Methods for the Behavioral Sciences. Rinehart & Company, Inc., Second Printing. 1955.
- Flippo, Edwin B. Principles of Personnel Management. New York: McGraw-Hill Book Company, Inc., 1961.
- Garrett, Henry E. Statistics in Psychology and Education. 5th ed. New York: Longmans, Green and Co., Reprinted, July 1960.
- Glanz, Edward C. Groups in Guidance. Boston: Allyn and Bacon, Inc., 1962.
- Goldman, Leo. Using Tests in Counseling. New York: Appleton-Century-Crofts, Inc., 1961.
- Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton-Mifflin, 1953.
- Lyman, Howard B. Test Scores and What They Mean. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963.

- Ohlsen, Merle M. "Interpretation of Test Scores," The Impact and Improvement of School Testing Programs, National Society for the Study of Education, Sixty-Second Yearbook, Part II. Chicago: The University of Chicago Press, 1963.
- Super, D. E., and Crites, J. O. Appraising Vocational Fitness. revised ed. New York: Harper & Row, 1962.
- Tyler, Leona E. The Work of the Counselor. 2d ed., New York: Appleton-Century-Crofts, Inc., 1961.
- Wrenn, Gilbert C. The Counselor in a Changing World. Washington, D.C.: American Personnel and Guidance Association, 1962.

#### Articles and Periodicals

- "A Criterion for Counseling," Psychological Monogram, Vol. LXXI, No. 15 (1957).
- Berdie, Ralph F. "Changes in Self-ratings as a Method of Evaluating Counseling," Journal of Counseling Psychology, Vol. I (1954), pp. 49-54.
- Dressel, Paul L. and Matteson, Ross W. "The Effects of Client Participation in Test Interpretation," Educational and Psychological Measurement, Vol. X (1950), pp. 693-706.
- "Ethical Standards-American Personnel and Guidance Association," The Personnel and Guidance Journal, Vol. XL, No. 2 (October, 1961), p. 207.
- Froehlich, C.P. "Does Test Taking Change Self Ratings?" California Journal of Educational Research, Vol. V (1954), pp. 166-169; 175.
- Froehlich, C.I. and Moser, W.E. "Do Counselees Remember Test Scores?", Journal of Counseling Psychology, Vol. I (1954), pp. 149-152.
- Gustad, J. W. and Tuma, A. H. "The Effects of Client and Counselor Personality Characteristics on Client Learning in Counseling," Journal of Counseling Psychology, Vol. IV (1957), pp. 136-143.
- \_\_\_\_\_. "The Effects of Different Methods of Test Introduction and Interpretation on Client Learning in Counseling," Journal of Counseling Psychology, Vol. IV, No. 4 (1957), pp. 313-317.

- Holmes, June Elaine. "The Comparisons of Four Techniques Used in Presenting Test Information to Freshmen Students," Dissertation Abstracts, Vol. XXI (1961), p. 3379.
- Hoyt, Donald P. "An Evaluation of Group and Individual Programs in Vocational Guidance," Journal of Applied Psychology, Vol. XXXIX (1955), pp. 26-30.
- Johnson, D. G. "Effect of Vocational Counseling on Self-knowledge," Educational and Psychological Measurement, Vol. XIII (1953), pp. 330-338.
- Lallas, J. E. "A Comparison of Three Methods of Interpretation of Achievement Tests to Students," Dissertation Abstracts, Vol. XVI (1956), p. 1842.
- Lane, David. "A Comparison of Two Techniques of Interpreting Test Results to Clients in Vocational Counseling," Dissertation Abstracts, Vol. XII (1952), pp. 591-592.
- O'Hara, Robert P. and Tiedeman, David J. "Vocational Self-concept in Adolescence," Journal of Counseling Psychology, Vol. VI (1959), pp. 292-301.
- Page, Ellis B. "Counseling With Mass Media," Letter to the Editor, The Personnel and Guidance Journal, Vol. XLI, No. 8 (April, 1963), p. 729.
- Parker, H. J. "A Study of Immediate Recall, Delayed Recall, and Distortion of Objective Data in Counseling," Dissertation Abstracts, Vol. XVII (1957), pp. 305-306.
- Robertson, M. H. "A Comparison of Counselor and Student Reports of Counseling Interviews," Journal of Counseling Psychology, Vol. V (1958), pp. 276-280.
- Rogers, L. B. "A Comparison of Two Kinds of Test Interpretation Interviews," Journal of Counseling Psychology, Vol. I (1954), pp. 224-231.
- Schulman, Jacob. "A Comparison Between Ninth and Twelfth Grade Students on Self-estimates of Abilities and Objective Scores on the Differential Aptitude Tests," Dissertation Abstracts, Vol. XVI (1956), pp. 285-286.
- Singer, S. L. and Stefflre, B. "Analysis of the Self-estimate in the Evaluation of Counseling," Journal of Counseling Psychology, Vol. I (1954), pp. 252-255.

- Snider, G. R. and Hunnicut, H. B. "Testing in Oklahoma Schools," Oklahoma Teacher, Vol. XLII (1961), pp. 12-17.
- Torrance, E. P. "Some Practical Uses of a Knowledge of Self-concepts in Counseling and Guidance," Educational and Psychological Measurements, Vol. XIV (1954), pp. 120-127.
- Womer, Frank B. "Testing Programs-Misconception, Misuse, Overuse," Michigan Journal of Secondary Education (Spring, 1961), p. 153.

### Reports

- Gustad, J. W. The Effects of Differing Methods of Test Selection and Test Interpretation on Learning in the Interview. Office of Naval Research, Contract No. 1225 (oo), 1955.
- Lister, James L. and Ohlsen, Merle M. The Effects of Orientation to Testing on Motivation for and Outcomes of Test Interpretation. Cooperative Research Project No. 1344, 1962.

### Unpublished Material

- Adamek, Edward G. "The Effects of Testing and Test Interpretation on Selected Self-perceptions," Unpublished Doctoral Thesis, University of Illinois, 1961.
- Lundy, Charles T. and Shertzer, Bruce. "The Relationship of D.A.T. Scores to High School Marks." Paper read before the American Personnel and Guidance Association National Convention, Boston, Massachusetts, April 8, 1963.
- Wright, E. Wayne. "A Comparison of Individual and Multiple Counseling for Test Interpretation Interviews," Unpublished Doctoral Dissertation, University of California, Berkeley, 1957.

APPENDIX A

STATISTICAL DATA



## LEGEND OF ABBREVIATIONS

SCAT(T)--School and College Ability Test--Total Raw Score

PT-Q--Pretest Questionnaire (Criterion Instrument)

IR-Q--Immediate Recall Questionnaire (Criterion Instrument)

IR(Gain)--Immediate Recall Gain Score

DR-Q--Delayed Recall Questionnaire (Criterion Instrument)

DR(Gain)--Delayed Recall Gain Score

PT-DE--Pretest, Direction of Error

IR-DE--Immediate Recall, Direction of Error

DR-DE--Delayed Recall, Direction of Error

INST-I--Special Marking Instruction--Recall--Control group  
only

INST-II--Special Marking Instruction--Acceptance--Control  
group only

RT-Q--Retest Questionnaire (Criterion Instrument)--Control  
group only

SRF--Student Report Folder--Published by Educational Testing  
Service for STEP/SCAT

TABLE 10

CELL # 1 REPORTING CONDITION GROUP ABILITY LEVEL L

Student No.	Sex	SCAT(T)	PT-Q	IR-Q	IR(Gain)	DR-Q	DR(Gain)	PT-DE	IR-DE	DR-DE
1	F	43	56	64	+ 8	64	+ 8	+ 9	+ 6	+ 5
2	M	46	68	66	- 2	72	+ 4	+ 6	+ 5	+ 1
3	M	40	54	56	+ 2	62	+ 8	+10	+11	+ 8
4	F	50	58	72	+14	66	+ 8	-10	- 4	- 7
5	M	46	56	74	+18	64	+ 8	+ 7	+ 3	+ 6
6	M	39	66	60	- 6	62	- 4	+ 5	+ 6	+ 7
7	F	46	70	78	+ 8	78	+ 8	+ 4	- 2	+ 1
8	M	49	64	78	+14	68	+ 4	- 4	- 3	- 5
9	M	29	66	72	+ 6	64	- 2	+ 2	- 2	- 3
10	F	39	70	66	- 4	76	+ 6	+ 3	- 6	+ 2
11	M	37	64	52	-12	54	-10	+ 6	+ 8	+ 8
12	F	43	66	76	+10	72	+ 6	- 6	- 4	- 6
13	M	15	54	62	+ 8	50	- 4	+ 1	+ 2	+ 2
14	F	50	62	68	+ 6	68	+ 6	+ 2	- 2	0
15	F	38	64	66	+ 2	68	+ 4	+ 4	+ 1	- 1
16	F	25	58	54	- 4	52	- 6	0	+ 1	+ 2
17	F	50	72	74	+ 2	74	+ 2	0	- 3	- 4
18	F	50	54	64	+10	58	+ 4	- 7	- 1	0
19	M	47	58	58	0	54	- 4	+ 5	+ 5	+ 2
20	M	42	60	72	+12	68	+ 8	- 6	- 4	- 3
Sum		824	1240	1332	92	1294	54	31	17	15
Mean		41.200	62.000	66.600	4.600	64.700	2.700	1.550	.850	.750
Variance		85.105	33.684	62.368	59.421	63.684	31.684			
Stand. Dev.		9.225	5.804	7.897	7.708	7.980	5.629			

TABLE 11

CELL # 2 REPORTING CONDITION GROUP ABILITY LEVEL M

Student No.	Sex	SCAT(T)	PT-Q	IR-Q	IR(Gain)	DR-Q	DR(Gain)	PT-DE	IR-DE	DR-DE
1	M	69	66	60	- 6	56	-10	- 5	- 6	- 7
2	M	70	62	76	+14	76	+14	+ 3	+ 4	+ 2
3	F	74	52	58	+ 6	58	+ 6	- 6	-10	- 7
4	M	65	54	62	+ 8	64	+10	-10	- 7	- 8
5	F	65	70	74	+ 4	66	- 4	- 1	+ 3	+ 4
6	M	61	60	74	+14	64	+ 4	- 9	- 3	- 6
7	F	70	56	56	0	60	+ 4	- 9	- 7	- 7
8	F	66	62	76	+14	70	+ 8	0	+ 3	- 3
9	F	62	70	74	+ 4	68	- 2	- 1	+ 1	- 1
10	M	59	66	74	+ 8	70	+ 4	- 6	- 3	- 3
11	M	72	60	72	+12	62	+ 2	- 9	- 3	- 7
12	F	60	60	60	0	60	0	+ 3	+ 9	+ 6
13	F	72	52	74	+22	68	+16	- 5	- 5	- 5
14	F	72	58	60	+ 2	56	- 2	- 6	- 5	- 7
15	M	64	64	68	+ 4	70	+ 6	- 4	- 2	- 5
16	F	66	46	70	+24	68	+22	- 2	0	- 3
17	M	73	60	70	+10	70	+10	- 3	- 3	+ 2
18	M	59	70	66	- 4	62	- 8	- 6	- 9	-10
19	M	63	50	66	+16	46	- 4	-14	- 9	-12
20	F	61	64	64	0	60	- 4	- 6	- 5	- 4
Sum		1323	1202	1354	152	1274	72	-96	-57	-81
Mean		66.150	60.100	67.700	7.600	63.700	3.600	-4.800	-2.850	-4.050
Variance		25.105	46.947	43.474	66.368	46.421	67.632			
Stand. Dev.		5.010	6.852	6.593	8.147	6.813	8.224			

TABLE 12

CELL # 3 REPORTING CONDITION GROUP ABILITY LEVEL H

Student No.	Sex	SCAT(T)	PT-Q	IR-Q	IR(Gain)	DR-Q	DR(Gain)	PT-DE	IR-DE	DR-DE
1	M	84	62	78	+16	62	0	- 8	- 3	- 7
2	F	95	74	72	- 2	60	-14	+ 1	- 4	- 8
3	M	93	74	84	+10	82	+ 8	- 2	0	- 1
4	F	89	42	64	+22	56	+14	- 6	- 5	- 6
5	F	87	82	80	- 2	82	0	- 1	0	- 1
6	M	83	64	80	+16	76	+12	+ 1	0	0
7	F	86	62	74	+12	76	+14	- 4	- 2	- 2
8	M	92	62	84	+22	82	+20	- 4	0	+ 1
9	F	89	68	80	+12	84	+16	- 4	+ 2	0
10	F	85	56	62	+ 6	56	0	- 6	- 6	- 6
11	M	94	58	74	+16	70	+12	- 7	- 2	- 3
12	F	89	54	72	+18	74	+20	- 7	- 4	- 1
13	M	90	50	80	+30	78	+28	- 7	0	- 1
14	M	87	70	82	+12	72	+ 2	+ 5	- 1	0
15	M	86	58	78	+20	76	+18	-11	- 2	- 4
16	F	92	50	78	+28	72	+22	- 8	+ 3	+ 2
17	F	91	70	70	0	62	- 8	- 5	- 5	- 9
18	M	85	40	72	+32	60	+20	- 8	0	- 6
19	F	86	62	62	0	66	+ 4	- 5	- 4	- 5
20	M	96	80	84	+ 4	80	0	+ 2	0	0
Sum		1779	1238	1510	272	1426	188	-91	-33	-57
Mean		88.950	61.900	75.500	13.600	71.300	9.400	-4.550	-1.650	-2.850
Variance		14.579	130.737	48.789	108.474	85.579	121.316			
Stand. Dev.		3.818	11.434	6.985	10.415	9.251	11.014			

TABLE 13

CELL # 4 REPORTING CONDITION INDIVIDUAL ABILITY LEVEL L

Student No.	Sex	SCAT(T)	PT-Q	IR-Q	IR(Gain)	DR-Q	DR(Gain)	PT-DE	IR-DE	DR-DE
1	F	41	70	76	+ 6	72	+ 2	+ 5	- 2	0
2	M	47	70	74	+ 4	74	+ 4	- 4	+ 5	0
3	M	35	40	58	+18	38	- 2	+11	+ 8	+12
4	M	36	48	52	+ 4	54	+ 6	+ 6	+ 3	+ 5
5	M	49	62	66	+ 4	60	- 2	- 4	- 3	- 3
6	F	33	66	66	0	66	0	+ 2	- 1	+ 1
7	F	47	56	62	+ 6	64	+ 8	- 1	+ 2	0
8	M	44	44	56	+12	58	+14	+ 8	+ 5	+ 6
9	F	43	58	74	+16	68	+10	0	- 3	- 7
10	M	41	62	62	0	64	+ 2	+ 5	+ 3	+ 6
11	F	28	54	64	+10	60	+ 6	+ 2	+ 1	+ 1
12	M	41	58	76	+18	74	+16	+ 4	- 4	- 3
13	F	46	68	62	- 6	68	0	+ 3	+ 6	+ 4
14	M	42	52	80	+28	74	+22	+ 8	0	+ 1
15	F	33	56	64	+ 8	56	0	+ 6	+ 5	+ 6
16	M	45	62	70	+ 8	64	+ 2	+ 1	- 1	+ 1
17	F	40	68	72	+ 4	62	- 6	- 2	- 2	- 2
18	F	33	70	64	- 6	68	- 2	- 1	- 2	- 1
19	M	36	62	68	+ 6	58	- 4	- 1	0	- 2
20	F	47	68	78	+10	72	+ 4	- 1	+ 3	+ 4
Sum		807	1194	1344	150	1274	80	47	23	29
Mean		40.350	59.700	67.200	7.500	63.700	4.000	2.350	1.150	1.450
Variance		35.105	78.000	129.158	62.053	75.053	50.526			
Stand. Dev.		5.925	8.832	11.364	7.877	8.663	7.108			

TABLE 14

CELL # 5 REPORTING CONDITION INDIVIDUAL ABILITY LEVEL M

Student No.	Sex	SCAT(T)	PT-Q	IR-Q	IR(Gain)	DR-Q	DR(Gain)	PT-DE	IR-DE	DR-DE
1	M	65	68	72	+ 4	68	0	+ 3	+ 5	+ 6
2	F	67	64	78	+14	68	+ 4	- 7	- 3	- 5
3	M	72	62	78	+16	70	+ 8	- 1	+ 1	- 1
4	F	67	64	68	+ 4	64	0	- 5	- 5	- 5
5	M	73	56	78	+22	56	0	+ 1	0	- 2
6	M	63	60	66	+ 6	58	- 2	+ 1	+ 1	+ 5
7	F	67	58	70	+12	72	+14	- 7	- 4	- 3
8	M	60	66	82	+16	80	+14	- 6	- 1	- 1
9	M	68	56	82	+26	66	+10	- 6	- 1	- 7
10	F	64	64	68	+ 4	62	- 2	- 3	- 3	- 5
11	F	71	72	76	+ 4	74	+ 2	- 2	- 2	- 1
12	M	67	44	60	+16	50	+ 6	-12	- 6	- 8
13	F	67	64	76	+12	72	+ 8	- 4	- 4	- 6
14	F	65	70	74	+ 4	80	+10	- 5	- 1	- 2
15	F	67	60	72	+12	60	0	- 3	+ 1	- 2
16	M	59	44	50	+ 6	56	+12	-13	-11	-11
17	M	68	62	76	+14	78	+16	- 9	0	+ 2
18	M	65	64	76	+12	70	+ 6	- 3	- 2	- 5
19	F	69	64	72	+ 8	72	+ 8	- 6	- 1	- 4
20	F	67	62	82	+20	68	+ 6	- 8	- 1	- 6
Sum		1331	1224	1456	232	1344	120	-95	-37	-58
Mean		66.550	61.200	72.800	11.600	67.200	6.000	-4.750	-1.850	-2.900
Variance		12.053	74.909	61.211	43.211	68.368	31.579			
Stand. Dev.		3.471	8.655	7.824	6.574	8.268	5.619			

TABLE 15

CELL # 6 REPORTING CONDITION INDIVIDUAL ABILITY LEVEL H

Student No.	Sex	SCAT(T)	PT-Q	IR-Q	IR(Gain)	DR-Q	DR(Gain)	PT-DE	IR-DE	DR-DE
1	F	101	52	84	+32	82	+30	- 8	0	- 1
2	F	95	46	70	+24	56	+10	- 6	- 4	- 3
3	F	86	70	76	+ 6	74	+ 4	0	- 4	- 5
4	M	85	60	76	+16	66	+ 6	- 7	- 2	- 4
5	F	89	52	78	+26	68	+16	- 7	- 3	- 6
6	F	86	52	64	+12	68	+16	- 3	- 2	- 6
7	F	93	50	60	+10	52	+ 2	- 4	- 2	- 4
8	M	88	74	76	+ 2	74	0	- 3	+ 4	+ 5
9	F	85	42	70	+28	60	+18	- 4	- 3	- 3
10	F	83	62	80	+18	68	+ 6	- 6	- 2	- 3
11	M	91	50	74	+24	74	+24	- 4	+ 3	+ 5
12	M	100	58	76	+18	76	+18	- 6	0	0
13	F	83	64	78	+14	78	+14	- 4	- 1	+ 1
14	F	91	68	68	0	74	+ 6	+ 1	+ 1	+ 1
15	M	96	62	82	+20	78	+16	- 6	+ 1	- 3
16	M	97	70	74	+ 4	76	+ 6	- 1	+ 1	+ 2
17	M	89	72	78	+ 6	64	- 8	- 3	- 3	- 6
18	M	84	66	74	+ 8	82	+16	- 3	- 2	- 1
19	M	92	46	54	+ 8	56	+10	- 7	- 7	- 7
20	M	<u>100</u>	<u>70</u>	<u>82</u>	<u>+12</u>	<u>82</u>	<u>+12</u>	<u>- 6</u>	<u>+ 1</u>	<u>+ 1</u>
Sum		1814	1186	1474	288	1408	222	-87	-24	-37
Mean		90.700	59.300	73.700	14.400	70.400	11.100	-4.350	-1.200	-1.850
Variance		34.632	98.211	57.368	84.053	82.789	76.421			
Stand. Dev.		5.885	9.910	7.574	9.168	9.099	8.742			

TABLE 16

CELL # 2 REPORTING CONDITION SRF ABILITY LEVEL L

Student No.	Sex	SCAT(T)	PT-Q	IR-Q	IR(Gain)	DR-Q	DR(Gain)	PT-DE	IR-DE	DR-DE
1	M	43	64	62	- 2	66	+ 2	- 1	+ 1	- 3
2	F	48	62	66	+ 4	54	- 8	- 4	- 4	- 6
3	M	42	58	64	+ 6	60	+ 2	+ 2	- 1	+ 1
4	F	19	74	70	- 4	70	- 4	- 1	- 4	- 4
5	F	50	72	66	- 6	66	- 6	- 3	- 5	- 5
6	M	46	70	76	+ 6	76	+ 6	- 5	- 2	0
7	M	26	56	54	- 2	54	- 2	+ 2	+ 3	+ 1
8	F	42	58	60	+ 2	62	+ 4	+ 8	+ 6	+ 6
9	M	50	72	72	0	72	0	- 4	- 2	0
10	F	47	66	68	+ 2	74	+ 8	0	+ 2	0
11	F	43	64	72	+ 8	60	- 4	- 6	- 6	- 2
12	M	36	74	72	- 2	64	-10	+ 1	+ 1	+ 3
13	F	37	66	64	- 2	60	- 6	+ 3	+ 2	+ 6
14	M	47	58	60	+ 2	66	+ 8	+ 5	0	- 1
15	M	46	56	64	+ 8	60	+ 4	- 3	0	+ 2
16	F	44	64	72	+ 8	68	+ 4	- 7	- 5	- 4
17	M	48	66	68	+ 2	64	- 2	+ 4	+ 3	- 2
18	M	35	50	70	+20	36	-14	- 8	- 2	0
19	F	49	68	66	- 2	66	- 2	+ 2	- 1	+ 2
20	F	37	68	72	+ 4	70	+ 2	- 3	0	- 1
Sum		835	1286	1338	52	1268	-18	-18	-14	- 7
Mean		41.750	64.300	66.900	2.600	63.400	-.900	-.900	-.700	-.350
Variance		66.105	44.316	28.842	34.158	76.684	36.000			
Stand. Dev.		8.130	6.657	5.370	5.844	8.757	6.000			



TABLE 17

CELL # 8 REPORTING CONDITION SRF ABILITY LEVEL M

Student No.	Sex	SCAT(T)	PT-Q	IR-Q	IR(Gain)	DR-Q	DR(Gain)	PT-DE	IR-DE	DR-DE
1	M	66	62	74	+12	62	0	+ 5	+ 5	+ 5
2	F	73	54	80	+26	74	+20	- 8	- 2	- 5
3	M	64	58	74	+16	68	+10	- 5	0	- 1
4	M	70	76	74	- 2	76	0	- 1	+ 5	+ 2
5	F	62	52	52	0	54	+ 2	+ 6	+ 1	+ 7
6	M	64	62	68	+ 6	70	+ 8	- 5	- 1	+ 5
7	M	73	52	62	+10	44	- 8	- 7	- 6	- 7
8	F	64	64	66	+ 2	56	- 8	- 4	- 4	- 5
9	F	67	68	78	+10	64	- 4	- 2	- 1	- 4
10	F	74	50	64	+14	64	+14	+ 5	+ 1	+ 6
11	M	62	72	76	+ 4	68	- 4	- 5	- 2	- 5
12	M	63	64	72	+ 8	64	0	- 4	- 1	- 4
13	M	64	50	74	+24	70	+20	- 3	- 1	0
14	M	73	54	60	+ 6	60	+ 6	0	0	0
15	M	59	58	80	+22	78	+20	- 8	+ 2	- 3
16	F	62	58	76	+18	64	+ 6	+ 9	- 4	- 8
17	F	69	62	58	- 4	58	- 4	- 7	- 6	- 5
18	F	65	64	66	+ 2	68	+ 4	0	- 2	- 1
19	F	64	58	72	+14	64	+ 6	- 6	- 2	- 3
20	F	61	64	74	+10	62	- 2	- 6	- 2	- 4
Sum		1319	1202	1400	198	1288	86	-64	-20.000	-30
Mean		65.950	60.100	70.000	9.900	64.400	4.300	-3.200	-1.000	-1.500
Variance		20.474	50.316	56.474	72.210	62.158	60.000			
Stand. Dev.		4.524	7.093	7.515	8.498	7.884	7.746			

TABLE 18

CELL # 9 REPORTING CONDITION SRF ABILITY LEVEL H

Student No.	Sex	SCAT(T)	PT-Q	IR-Q	IR(Gain)	DR-Q	DR(Gain)	PT-DE	IR-DE	DR-DE
1	F	96	74	74	0	76	+ 2	+ 4	+ 5	+ 1
2	M	83	52	72	+20	64	+12	- 6	- 2	- 6
3	F	84	58	78	+20	78	+20	- 6	+ 1	+ 2
4	F	84	60	76	+16	78	+18	- 8	+ 4	+ 1
5	M	90	62	78	+16	72	+10	- 7	- 3	- 6
6	M	89	54	56	+ 2	52	- 2	- 6	- 5	- 4
7	M	86	62	74	+12	62	0	- 8	- 3	- 9
8	F	90	70	76	+ 6	76	+ 6	- 7	- 2	- 4
9	M	84	62	70	+ 8	60	- 2	- 6	- 3	- 5
10	M	85	54	74	+20	58	+ 4	- 7	- 1	- 6
11	M	83	56	78	+22	58	+ 2	- 6	- 1	- 4
12	F	86	62	72	+10	74	+12	- 8	- 5	- 5
13	F	84	52	80	+28	66	+14	- 5	- 2	- 6
14	M	98	62	70	+ 8	66	+ 4	- 7	- 7	- 8
15	F	93	52	82	+30	76	+24	- 6	- 1	- 2
16	M	88	56	82	+26	70	+14	- 6	+ 1	0
17	M	100	72	72	0	74	+ 2	+ 5	+ 5	+ 4
18	F	92	62	82	+20	76	+14	- 7	+ 1	+ 4
19	F	97	52	72	+20	70	+18	- 4	0	0
20	F	93	60	78	+18	70	+10	- 7	- 3	- 3
Sum		1785	1194	1496	302	1376	182	-108	-21	-56
Mean		89.250	59.700	74.800	15.100	68.800	9.100	-5.400	-1.050	-2.800
Variance		29.684	43.474	34.684	80.632	59.105	59.579			
Stand. Dev.		5.448	6.593	5.889	8.979	7.688	7.719			

TABLE. 19

CELL # 10 REPORTING CONDITION CONTROL ABILITY LEVEL L

Student No.	Sex	SCAT(T)	PT-Q	RT-Q	RT(Gain)	IR-Q	IR(Gain)	INST-I	INST-II
1	M	50	76	66	-10	58	-18	58	50
2	M	47	64	58	-6	68	+4	66	72
3	M	32	56	60	+4	50	-6	58	38
4	M	40	66	68	+2	70	+4	76	60
5	M	46	68	56	-12	64	-4	70	68
6	F	50	56	56	0	66	+10	64	64
7	M	34	52	56	+4	54	+2	62	46
8	M	46	54	58	+4	62	+8	64	64
9	M	46	60	54	-6	48	-12	50	42
10	F	50	68	68	0	64	-4	78	70
11	F	46	64	64	0	74	+10	68	74
12	M	41	66	68	+2	64	-2	72	68
13	M	26	64	58	-6	48	-16	58	56
14	M	48	70	66	-4	72	+2	70	70
15	F	46	64	68	+4	66	+2	68	66
16	F	45	70	60	-10	66	-4	66	66
17	F	36	60	48	-12	68	+8	74	70
18	F	42	70	58	-12	72	+2	72	66
19	F	48	68	70	+2	76	+8	74	68
20	F	37	60	66	+6	72	+12	70	68
Sum		856	1276	1226	-50	1282	6	1338	1246
Mean		42.800	63.800	61.300	-2.500	64.100	.300	66.900	62.300
Variance		44.789	38.263	36.316	39.105	71.789	73.368	50.316	107.895
Stand. Dev.		6.692	6.186	6.026	6.253	8.473	8.566	7.093	10.387

TABLE 20

CELL # 11 REPORTING CONDITION CONTROL ABILITY LEVEL M

Student No.	Sex	SCAT(T)	PT-Q	RT-Q	RT(Gain)	IR-Q	IR(Gain)	INST-I	INST-II
1	M	70	54	54	0	74	+20	72	70
2	F	69	62	60	- 2	72	+10	74	70
3	M	62	52	64	+12	60	+ 8	76	62
4	M	69	46	52	+ 6	70	+24	74	50
5	F	73	68	62	- 6	70	+ 2	72	74
6	F	65	70	64	- 6	70	0	68	74
7	F	72	60	64	+ 4	76	+16	78	66
8	M	68	44	64	+20	66	+22	74	76
9	M	72	50	56	+ 6	66	+16	78	58
10	F	71	54	44	-10	56	+ 2	64	68
11	F	60	68	60	- 8	64	- 4	72	70
12	M	71	58	68	+10	72	+14	74	66
13	F	61	62	72	+10	74	+12	76	74
14	F	63	64	66	+ 2	68	+ 4	82	70
15	M	65	58	58	0	64	+ 6	62	64
16	M	74	50	56	+ 6	68	+18	70	68
17	F	71	48	50	+ 2	56	+ 8	74	64
18	M	72	66	60	- 6	78	+12	80	76
19	F	60	60	56	- 4	70	+10	66	54
20	M	62	66	72	+ 6	72	+ 6	78	74
Sum		1350	1160	1202	42	1366	206	1464	1348
Mean		67.500	58.000	60.100	2.100	68.300	10.300	73.200	67.400
Variance		22.579	62.316	50.737	58.737	36.316	57.789	27.105	51.421
Stand. Dev.		4.751	7.894	7.123	7.664	6.026	7.602	5.206	7.171

TABLE 21

CELL # 12 REPORTING CONDITION CONTROL ABILITY LEVEL H

Student No.	Sex	SCAT(T)	PT-Q	RT-Q	RT(Gain)	IR-Q	IR(Gain)	INST-I	INST-II
1	F	87	54	58	+ 4	54	0	76	62
2	M	83	52	52	0	70	+18	78	74
3	M	89	60	64	+ 4	66	+ 6	70	64
4	M	92	62	76	+14	82	+20	82	82
5	F	89	62	70	+ 8	82	+20	82	82
6	F	87	50	52	+ 2	68	+18	68	64
7	M	99	72	72	0	72	0	74	74
8	M	86	54	56	+ 2	74	+20	76	70
9	M	89	64	72	+ 8	78	+14	82	78
10	M	94	60	58	- 2	72	+12	72	72
11	F	91	58	60	+ 2	76	+18	82	76
12	F	91	38	48	+10	68	+30	72	60
13	F	84	50	54	+ 4	60	+10	70	68
14	F	94	70	70	0	84	+14	84	84
15	M	91	68	72	+ 4	82	+14	78	80
16	F	84	68	68	0	74	+ 6	74	72
17	M	87	58	56	- 2	72	+14	74	68
18	M	89	50	42	- 8	66	+16	60	60
19	F	97	80	78	- 2	78	- 2	78	78
20	F	102	84	84	0	84	0	84	84
Sum		1805	1214	1262	48	1462	248	1516	1452
Mean		90.250	60.700	63.100	2.400	73.100	12.400	75.800	72.600
Variance		25.263	120.526	123.579	24.263	65.052	71.421	37.842	62.789
Stand. Dev.		5.026	10.978	10.902	4.925	8.065	8.457	6.152	7.924

TABLE 22

## DISTRIBUTION OF SCAT(T) RAW SCORES - GRADE TEN, EXPERIMENTAL SCHOOL

Experimental School						National Norm	
Raw Score Interval	Raw Score Mid-Point	Converted Score	Frequency	%ile	T Score	%ile	T Score
100-104	102	320	8	99.5	76	99.9	81
95-99	97	314	16	97.9	70	99.5	76
90-94	92	309	34	94.6	66	98	71
85-89	87	305	42	89.5	63	96	68
80-84	82	301	68	82.2	59	93	65
75-79	77	297	79	72.4	56	88	62
70-74	72	294	76	62.0	53	85	60
65-69	67	291	102	50.0	50	76	57
60-64	62	287	87	37.5	47	66	54
55-59	57	284	72	26.9	44	60	53
50-54	52	280	57	18.3	41	48	50
45-49	47	277	38	12.0	38	37	47
40-44	42	274	31	7.4	35	31	45
35-39	37	270	17	4.2	33	22	42
30-34	32	266	11	2.3	30	15	40
25-29	27	261	6	1.2	27	8	36
20-24	22	256	4	0.5	24	5	34
15-19	17	252	2	0.1	20	2	30

 $\bar{X} = 66.574$ 

Mdn = 66.950

S. D. = 16.290

Sk = -.069

Low Ability - Raw Score 50 or lower (n=117)

Middle Ability - Raw Scores 59 through 75 (n=212)

High Ability - Raw Score 83 or higher (n=120)

TABLE 23

SUMMARY BARTLETT'S TEST OF HOMOGENEITY OF VARIANCE  
BY ABILITY LEVEL--SCAT (T) RAW SCORES

Cell Number	df	$\chi^2$	$s^2$	$\log s^2$	$\underline{\chi^2}$
Low Ability Groups					
1	19	35566	85.105	1.9300	
4	19	33226	35.105	1.5454	
7	19	36117	66.105	1.8203	
10	19	37488	44.789	1.6512	
Totals			231.104	6.9469	4.3881
Middle Ability Groups					
2	19	87993	25.105	1.3999	
5	19	88807	12.053	1.0813	
8	19	87377	20.474	1.3113	
11	19	91554	22.579	1.3537	
Totals			80.211	5.1462	2.721
High Ability Groups					
3	19	158519	14.579	1.1638	
6	19	165188	34.632	1.5395	
9	19	159875	29.684	1.4725	
12	19	163381	25.263	1.4026	
Totals			104.158	5.5784	3.6924

TABLE 24

SUMMARY ANALYSIS OF VARIANCE BY ABILITY  
LEVEL--SCAT (T) RAW SCORES

Source of Variation	df	Sum of Squares	Mean Square (Variance)	F
Low Ability Groups				
Between the means of Conditions	3	63	21.000	2.745
Within Conditions	<u>76</u>	<u>4381</u>	57.645	
Total	79	4444		
Middle Ability Groups				
Between the means of Conditions	3	29	9.666	2.073
Within Conditions	<u>76</u>	<u>1523</u>	20.039	
Total	79	1552		
High Ability Groups				
Between the means of Conditions	3	40	13.333	1.960
Within Conditions	<u>76</u>	<u>1979</u>	26.039	
Total	79	2019		



TABLE 25

SCAT (T) RAW SCORE MEANS AND VARIANCES BY ABILITY LEVEL AND REPORTING CONDITION

Reporting Condition	Ability Level		
	Low	Middle	High
Group	$\bar{X}$ 41.200	$\bar{X}$ 66.150	$\bar{X}$ 88.950
	$s^2$ 85.105	$s^2$ 25.105	$s^2$ 14.579
	n=20	n=20	n=20
Individual	$\bar{X}$ 40.350	$\bar{X}$ 66.550	$\bar{X}$ 90.700
	$s^2$ 35.105	$s^2$ 12.053	$s^2$ 34.632
	n=20	n=20	n=20
SRF	$\bar{X}$ 41.750	$\bar{X}$ 65.950	$\bar{X}$ 89.250
	$s^2$ 66.105	$s^2$ 20.474	$s^2$ 29.684
	n=20	n=20	n=20
Control	$\bar{X}$ 42.800	$\bar{X}$ 67.500	$\bar{X}$ 90.250
	$s^2$ 44.789	$s^2$ 22.579	$s^2$ 25.263
	n=20	n=20	n=20

TABLE 26

PRE-TEST QUESTIONNAIRE MEANS AND VARIANCES BY ABILITY LEVEL AND REPORTING CONDITION

Reporting Condition	Ability Level		
	Low	Middle	High
Group	$\bar{X}$ 62.000	$\bar{X}$ 60.100	$\bar{X}$ 61.900
	$s^2$ 33.684	$s^2$ 46.947	$s^2$ 130.737
	n=20	n=20	n=20
Individual	$\bar{X}$ 59.700	$\bar{X}$ 61.200	$\bar{X}$ 59.300
	$s^2$ 78.000	$s^2$ 74.909	$s^2$ 98.211
	n=20	n=20	n=20
SRF	$\bar{X}$ 64.300	$\bar{X}$ 60.100	$\bar{X}$ 59.700
	$s^2$ 44.316	$s^2$ 50.316	$s^2$ 43.474
	n=20	n=20	n=20
Control	$\bar{X}$ 63.800	$\bar{X}$ 58.000	$\bar{X}$ 60.700
	$s^2$ 38.263	$s^2$ 62.316	$s^2$ 25.263
	n=20	n=20	n=20

TABLE 27

IMMEDIATE RECALL(IR) MEANS AND VARIANCES OF GAIN SCORES  
BY REPORTING CONDITION AND ABILITY LEVEL

Reporting Condition	Ability Level		
	Low	Middle	High
Group	$\bar{X}$ 4.600	$\bar{X}$ 7.600	$\bar{X}$ 13.600
	$s^2$ 59.421	$s^2$ 66.368	$s^2$ 108.474
	n=20	n=20	n=20
Individual	$\bar{X}$ 7.500	$\bar{X}$ 11.600	$\bar{X}$ 14.400
	$s^2$ 62.053	$s^2$ 43.211	$s^2$ 84.053
	n=20	n=20	n=20
SRF	$\bar{X}$ 2.600	$\bar{X}$ 9.900	$\bar{X}$ 15.100
	$s^2$ 34.158	$s^2$ 72.210	$s^2$ 84.053
	n=20	n=20	n=20
Control	$\bar{X}$ -2.500	$\bar{X}$ 2.100	$\bar{X}$ 2.400
	$s^2$ 39.105	$s^2$ 58.737	$s^2$ 24.263
	n=20	n=20	n=20

TABLE 28

CONSOLIDATION OF IMMEDIATE RECALL (IR-GAIN)  
DATA BY ABILITY LEVEL

	Ability Level		
	Low	Middle	High
Sum of Gain Scores	294	582	862
Mean Gain	4.900	9.700	14.367
Variance	88.464	61.264	88.338
	n=60	n=60	n=60

TABLE 29

CONSOLIDATION OF DELAYED RECALL (DR-GAIN) BY ABILITY LEVEL

	Ability Level		
	Low	Middle	High
Sum of Gain Scores	116	278	592
Mean Gain	1.933	4.633	9.867
Variance	42.436	58.372	83.643
	n=60	n=60	n=60

TABLE 30

DELAYED RECALL(DR) MEANS AND VARIANCES OF GAIN SCORES  
BY REPORTING CONDITION AND ABILITY LEVEL

Reporting Condition	Ability Level					
	Low		Middle		High	
Group	$\bar{X}$	2.700	$\bar{X}$	3.600	$\bar{X}$	9.400
	$s^2$	31.684	$s^2$	67.632	$s^2$	121.316
		n=20		n=20		n=20
Individual	$\bar{X}$	4.000	$\bar{X}$	6.000	$\bar{X}$	11.100
	$s^2$	50.526	$s^2$	31.579	$s^2$	76.421
		n=20		n=20		n=20
SRF	$\bar{X}$	-1.9900	$\bar{X}$	4.300	$\bar{X}$	9.100
	$s^2$	36.000	$s^2$	60.000	$s^2$	59.579
		n=20		n=20		n=20

TABLE 31

MEANS, VARIANCES, AND STANDARD DEVIATIONS OF GAIN SCORES FOR GROUP REPORTING OF  
TEST RESULTS TO CONTROL GROUPS USING THREE DIFFERENT MARKING INSTRUCTIONS

Marking Instruction	Ability Level		
	Low (n=20)	Middle (n=20)	High (n=20)
IR-Q	$\bar{X}$ .300	$\bar{X}$ 10.300	$\bar{X}$ 12.400
	$s^2$ 73.368	$s^2$ 57.789	$s^2$ 71.421
	s 8.566	s 7.602	s 8.457
INST-I	$\bar{X}$ 3.100	$\bar{X}$ 14.900	$\bar{X}$ 15.100
	$s^2$ 61.263	$s^2$ 83.158	$s^2$ 84.421
	s 7.827	s 9.119	s 9.188
INST-II	$\bar{X}$ -1.500	$\bar{X}$ 9.400	$\bar{X}$ 12.400
	$s^2$ 102.684	$s^2$ 55.632	$s^2$ 50.368
	s 10.133	s 7.459	s 7.097

TABLE 32

INTERCORRELATIONS BETWEEN FIVE DIFFERENT MARKINGS OF THE  
CRITERION INSTRUMENT BY THE CONTROL GROUP(n=60)

	Instrument Marking				
	PT-Q	RT-Q	IR-Q	INST-I	INST-II
PT-Q		.696*	.361*	.171	.379*
RT-Q			.528*	.488*	.493*
IR-Q				.733*	.772*
INST-I					.725*

\*significant at .01 level

TABLE 33

SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEANS OF GAIN SCORES WHEN THREE DIFFERENT MARKING INSTRUCTIONS WERE USED--CONTROL GROUP ONLY

Instruction	Mean Gain		Mean Gain <sub>diff</sub>	SE <sub>diff</sub>	t
Low Ability n=20					
IR-Q vs INST-I	.300	3.100	2.800	1.175	2.382**
IR-Q vs INST-II	.300	-1.500	1.800	1.250	1.440
INST-I vs INST-II	3.100	-1.500	4.600	1.467	3.135*
Middle Ability n=20					
IR-Q vs INST-I	10.300	14.900	4.600	2.084	2.207**
IR-Q vs INST-II	10.300	9.400	.900	2.060	.436
INST-I vs INST-II	14.900	9.400	5.500	2.379	2.312**
High Ability n=20					
IR-Q vs INST-I	12.400	15.100	2.700	1.558	1.732
IR-Q vs INST-II	12.400	12.400	.000	1.134	.000
INST-I vs INST-II	15.100	12.400	2.700	.750	3.600*
Total Group n=60					
IR-Q vs INST-I	7.667	11.033	3.366	.944	3.566*
IR-Q vs INST-II	7.667	6.767	.900	.868	1.036
INST-I vs INST-II	11.033	6.767	4.266	.979	4.358*

\*\*Significant at .05 level

\*Significant at .01 level



TABLE 34

DIRECTION AND MAGNITUDE OF ERRORS IN SELF-ESTIMATES BY ABILITY LEVEL AND  
MARKING INSTRUCTION--CONTROL GROUP ONLY

Subject Number	Ability Levels								
	Low			Middle			High		
	IR-Q	INST-I	INST-II	IR-Q	INST-I	INST-II	IR-Q	INST-I	INST-II
1	-13	-13	+17	+3	+2	-1	-13	-4	-11
2	+2	-1	+2	-4	-3	-1	-3	-3	-1
3	+17	+7	+21	-12	-4	-11	+3	+5	+4
4	+5	+2	+10	-5	-3	-11	+1	+1	+1
5	+2	+5	+2	-3	+4	+3	-1	-1	-1
6	+1	+2	+2	-1	-6	-3	-2	-2	-4
7	+11	+7	+17	+2	+3	+5	+6	+5	+5
8	-11	-8	-8	-8	-1	-2	-3	-4	-7
9	+16	+15	+19	-5	-3	-13	-1	+1	-1
10	+4	-1	+3	-8	+2	0	+6	+6	+6
11	-3	+2	+1	-2	0	+1	+2	+1	-4
12	-4	-2	-2	+2	+3	+5	-8	-6	-12
13	+10	+3	+4	-5	-4	-5	-2	+3	0
14	+4	-1	+3	-8	+1	-7	0	0	0
15	+1	-4	+1	-6	-7	-8	+1	+3	+2
16	-1	+5	+3	-6	-7	-4	-1	-1	+4
17	+4	+1	+2	-14	-4	-10	+2	+3	+3
18	0	-2	+5	-1	-1	-4	-7	-6	-6
19	-4	-3	+1	+1	+3	-8	+1	+1	+1
20	-2	-1	+2	+4	-3	+1	0	0	0
Sum	39	13	105	-76	-28	-73	-19	2	-21
$\bar{X}$	1.950	.650	5.250	-3.800	-1.400	-3.650	-.950	.100	-1.050

APPENDIX B

INSTRUMENTS AND MATERIALS

## QUESTIONNAIRE

NAME \_\_\_\_\_ GRADE \_\_\_\_\_  
                     Last                      First

SCHOOL \_\_\_\_\_ DATE \_\_\_\_\_

## INSTRUCTIONS

It is important that each person know and understand his or her achievement in certain subjects. Each person also needs to know the extent of his or her ability to learn in certain areas. Understanding of your Achievement and Ability will be of value to you in making your future educational and vocational plans. Filling out this form will help your counselor learn how well you understand your level of achievement and your ability to do school work. You need to know the meaning of the following terms so that you can mark this form accurately.

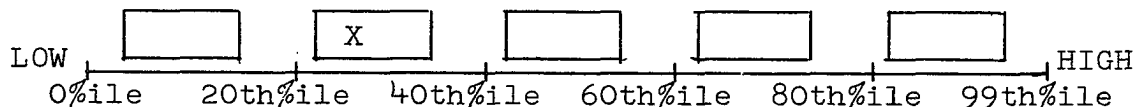
ABILITY: A measure of aptitude or capacity for learning.  
 (How much can I learn?)

ACHIEVEMENT: A measure of mastery of some particular subject area.  
 (How much have I learned?)

PERCENTILE (%ile) RANK: Indicates the percentage of students who have scores equal to or lower than yours. For example, a percentile rank of 70th percentile shows that 70% of the students had scores the same as yours or lower and that about 30% had higher scores.

The series of statements listed on the following pages are to be marked by placing an (X) inside the appropriate box along the line provided. For example, if you believe that you can throw a ball farther than about 35 out of 100 of your classmates (35%) then you would place your X in the second box from the left (between 20th%ile and 40th%ile) as shown below.

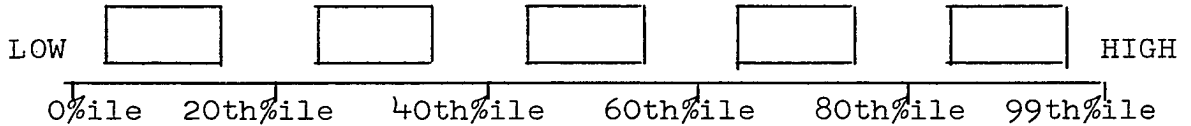
Example: I believe that my achievement in BALL-THROWING when compared with other students in my grade is about:



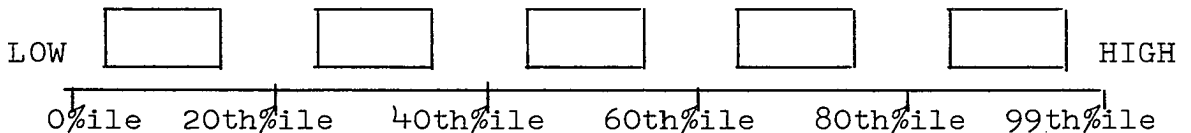
Your mark should appear inside one of the five boxes.

DO NOT TURN THE PAGE UNTIL INSTRUCTED TO DO SO

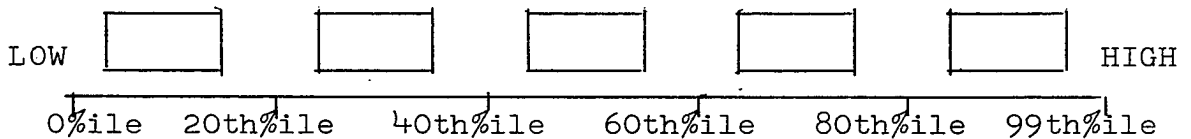
1. I believe that my achievement in the subject of MATHEMATICS when compared with other students in my grade is about:



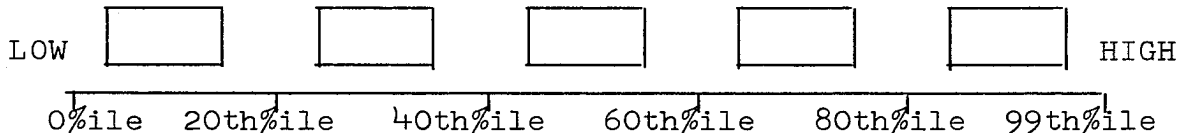
2. I believe that my achievement in the subject of SCIENCE when compared with other students in my grade is about:



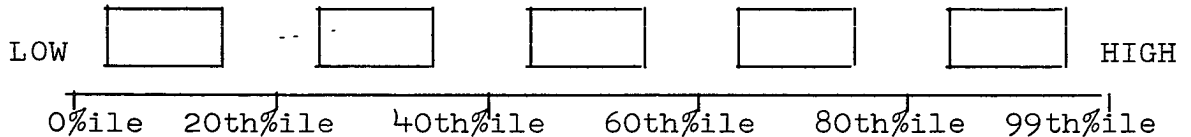
3. I believe that my achievement in the subject of SOCIAL STUDIES when compared with other students in my grade is about:



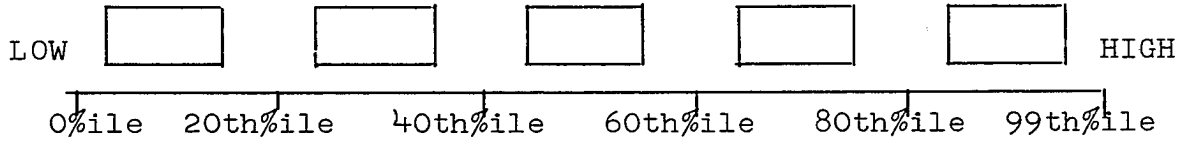
4. I believe that my achievement in the skill of READING when compared with other students in my grade is about:



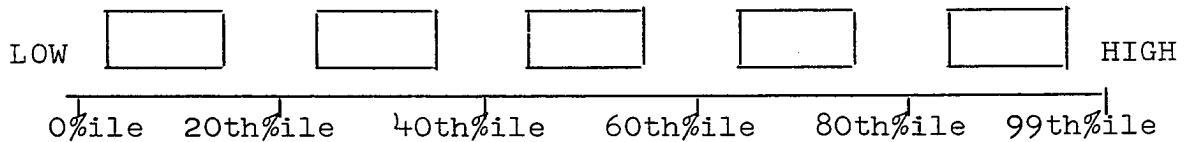
5. I believe that my achievement in the skill of WRITING (skill in accurately criticizing materials written by other people in terms of the ways they are organized and written) when compared with other students in my grade is about:



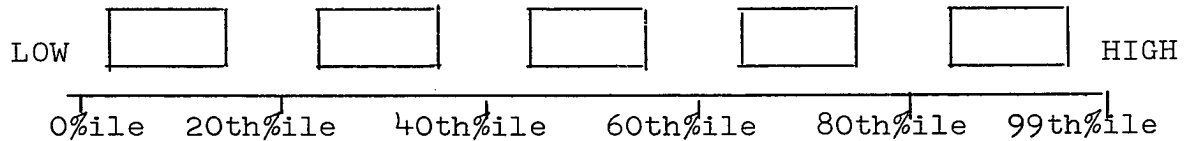
6. I believe that my VERBAL ABILITY (the ability to understand sentences and give the meanings of words) when compared with other students in my grade is about:



7. I believe that my QUANTITATIVE ABILITY (the ability to perform operations with numbers and to solve mathematics problems) when compared with other students in my grade is about:

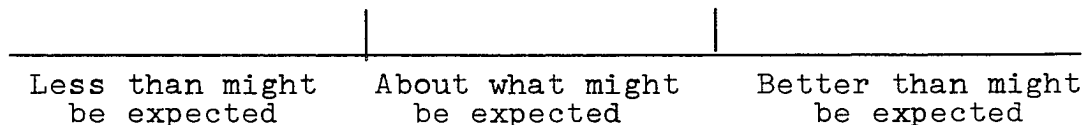


8. I believe that my OVERALL ABILITY to do school work both verbal and quantitative when compared with other students in my grade is about:

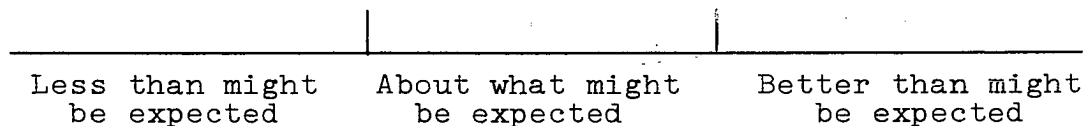


On the remaining five statements mark an (X) on the line above the words that best completes the statement for you.

9. When I compare my QUANTITATIVE ABILITY with my level of achievement in MATHEMATICS, I believe I am doing:



10. When I compare my OVERALL ABILITY to do school work with my level of achievement in SCIENCE, I believe I am doing:



11. When I compare my OVERALL ABILITY to do school work with my level of achievement in SOCIAL STUDIES, I believe I am doing:

---

Less than might  
be expected

About what might  
be expected

Better than might  
be expected

12. When I compare my VERBAL ABILITY with my level of achievement in READING, I believe I am doing:

---

Less than might  
be expected

About what might  
be expected

Better than might  
be expected

13. When I compare my OVERALL ABILITY to do school work with my level of achievement in WRITING (skill in accurately criticizing materials written by other people in terms of the ways they are organized and written) I believe I am doing:

---

Less than might  
be expected

About what might  
be expected

Better than might  
be expected

Procedure for Scoring the Criterion Instrument

The criterion instrument was designed to measure the relative congruence between student self-estimates and test-estimates following test interpretation. It was specifically designed to be used with the Sequential Tests of Educational Progress and the School and College Ability Test, (STEP/SCAT). Each of the first eight statements were marked along a continuum divided into five segments (quintiles). The student was asked to mark his position on the continuum in five achievement areas and in verbal, quantitative, and overall (total) ability. The last five statements required the student to indicate his position on a line comparing his achievement with his ability to do work in the specified areas.

The instrument was scored against the students actual scores on STEP/SCAT. In the first eight items eight(8) points were awarded for a mark in the same quintile as the actual score, six(6) points for a mark in an adjacent quintile, four(4) points for a mark two quintiles from the obtained score, two(2) points for a mark separated by three quintiles, and zero(0) for any mark separated from the obtained score by more than three quintiles. On the last five items a score of four(4) was awarded for a mark that accurately denotes the situation indicated by the profile regarding significance or non-significance of difference between ability and achievement, two(2) points were scored if the student marked an

adjacent segment, and zero(0) if the mark was separated from the correct interpretation by an empty space, i.e. marking significant underachievement when the profile actually indicates significant overachievement. On this scoring system the potential range of scores was from 0 to 84 points.

#### Scoring Procedure for Direction of Error

A score indicating the direction of error was obtained by adding algebraically symbols which were assigned to the thirteen items of the self-estimate questionnaire in the following manner: A plus(+) was assigned for each item on which the subject overestimated his position on the rating scale as compared with his test-estimate obtained from STEP/SCAT; a minus(-) was assigned for each item underestimated; and, a zero(0) was assigned when a given self-estimate coincided with the test-estimate for the same area.

#### Scoring Procedure for Magnitude and Direction of Error (Control Group Only)

A score indicating both direction and magnitude of error was obtained for each member of the control group in the following manner: A number and symbol was assigned to each of the thirteen items according to its deviation from the test-estimate. Plus numbers indicated overestimations, minus numbers indicated underestimations. The numbers increased arithmetically by quintiles as the magnitude of the error increased, i.e. an overestimation of one quintile was assigned a value of +1, an underestimation of two quintiles



was assigned a value of -2, etc. When self-estimate and test-estimate coincided, a value of 0 was assigned. The values thus assigned to the thirteen items were added algebraically to provide a score of magnitude and direction of error.

STEP/SCAT TEST INTERPRETATION CRITERION  
INSTRUMENT SCORE SHEET

Name \_\_\_\_\_ Date \_\_\_\_\_

School \_\_\_\_\_ Counselor \_\_\_\_\_

Condition \_\_\_\_\_ Level \_\_\_\_\_ Cell # \_\_\_\_\_  
(G, I or R) (L, M or H)

Pretest	Actual Scores (Quintiles)	I-Recall	D-Recall
1. _____	1. _____	1. _____	1. _____
2. _____	2. _____	2. _____	2. _____
3. _____	3. _____	3. _____	3. _____
4. _____	4. _____	4. _____	4. _____
5. _____	5. _____	5. _____	5. _____
6. _____	6. _____	6. _____	6. _____
7. _____	7. _____	7. _____	7. _____
8. _____	8. _____	8. _____	8. _____
	(1, 2 or 3)		
9. _____	9. _____	9. _____	9. _____
10. _____	10. _____	10. _____	10. _____
11. _____	11. _____	11. _____	11. _____
12. _____	12. _____	12. _____	12. _____
13. _____	13. _____	13. _____	13. _____
Total Score _____		Total Score _____	Total Score _____
		Total Gain _____	Total Gain _____

## STUDENT REPORT

\_\_\_\_\_  
(Student's Name)

**THE SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS (STEP)** are achievement tests designed to help you keep track of your own individual development in school and college. They will also help your teachers guide you in learning the important skills and processes that you go to school to learn. There are seven kinds of STEP tests: Mathematics, Science, Social Studies, Reading, Listening, Writing, and Essay Writing.

*Some facts you might like to know about STEP:*

- STEP tests measure how well you are able to *use* what you have learned and not just how much you remember. If you understand what you have learned about a subject, you should be able to use that learning in solving new problems in that field.
- The skills measured by STEP tests are skills that can be improved if you work at improving them.
- There are STEP tests suitable for students at dif-

ferent grade levels from fourth grade through the sophomore year of college.

- The problems in each test were chosen to be interesting to you and to measure how well you have learned the skills that most teachers think are important at your grade level.
- These tests try to measure how *well* you can work—not how *fast* you can work. Almost all students will be able to finish the tests for their grade in the time allowed.

**THE SCHOOL AND COLLEGE ABILITY TESTS (SCAT)** were designed to help you and your teachers estimate your general ability to do the work of your next school year. They do this by measuring your ability to use and understand words (verbal ability) and your ability to use and understand numbers (quantitative ability).

*Some facts you might like to know about SCAT:*

- SCAT tests measure your verbal and quantitative abilities because these are the two abilities you need most in order to succeed in school work.
- There are SCAT tests suitable for students at dif-

ferent grade levels from fourth grade through the sophomore year of college.

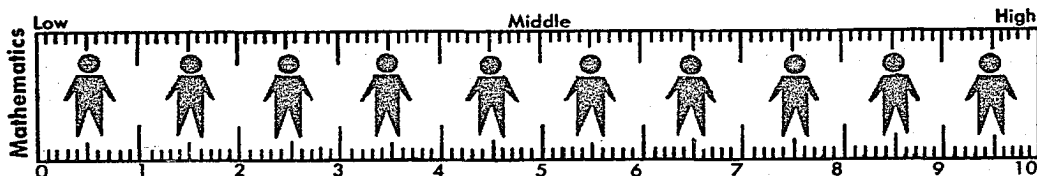
- Like STEP, the SCAT tests are not speed tests. Most students can complete each of the parts in the test for their grade during the time allowed.

Although this Report form provides space for recording results on SCAT and all STEP tests, your own school may not give all of them at any one time. Therefore some of the spaces may be blank.

These tests are part of a carefully planned educational program. The results are for your own benefit. They will help you see what your strengths and weaknesses are. Once you know these, you can decide more wisely what subjects you should work harder on and what courses of study will help you develop your own special talents and abilities.

On the next three pages of this folder are your test results.

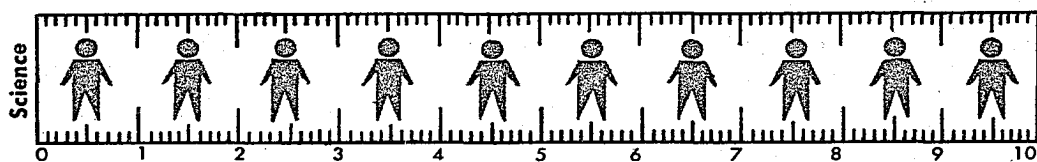
\_\_\_\_\_ 's standing on each of the tests taken is indicated by the shaded area.  
 (name)  
 Ability, as measured by each test, is compared with that of a sample of students in grade \_\_\_\_\_ in the  
 (grade)  
 \_\_\_\_\_  
 (nation or local group)



In the kind of ability measured by \_\_\_\_\_

about \_\_\_\_\_ out of 10 students are lower

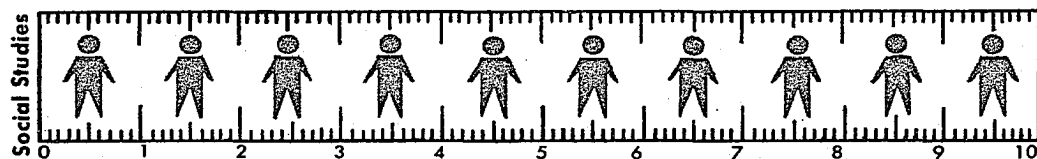
about \_\_\_\_\_ out of 10 students are higher



In the kind of ability measured by \_\_\_\_\_

about \_\_\_\_\_ out of 10 students are lower

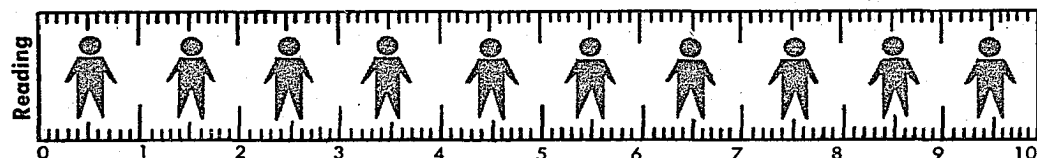
about \_\_\_\_\_ out of 10 students are higher



In the kind of ability measured by \_\_\_\_\_

about \_\_\_\_\_ out of 10 students are lower

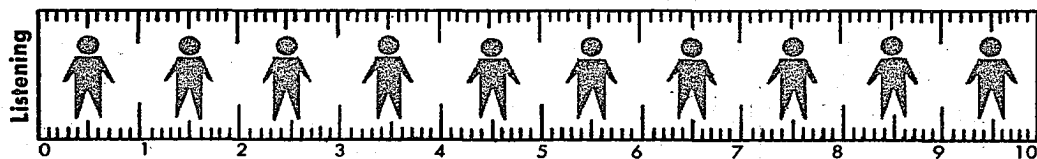
about \_\_\_\_\_ out of 10 students are higher



In the kind of ability measured by \_\_\_\_\_

about \_\_\_\_\_ out of 10 students are lower

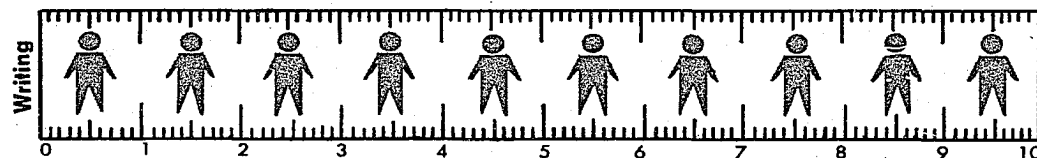
about \_\_\_\_\_ out of 10 students are higher



In the kind of ability measured by \_\_\_\_\_

about \_\_\_\_\_ out of 10 students are lower

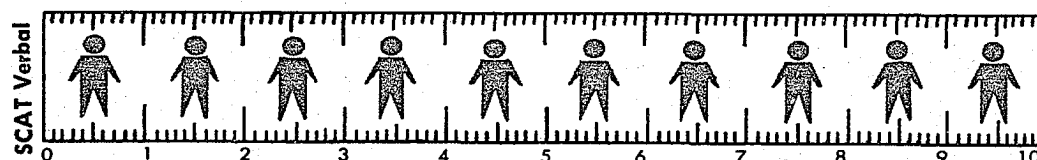
about \_\_\_\_\_ out of 10 students are higher



In the kind of ability measured by \_\_\_\_\_

about \_\_\_\_\_ out of 10 students are lower

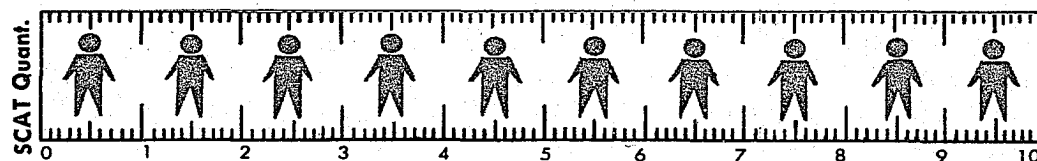
about \_\_\_\_\_ out of 10 students are higher



In the kind of ability measured by \_\_\_\_\_

about \_\_\_\_\_ out of 10 students are low

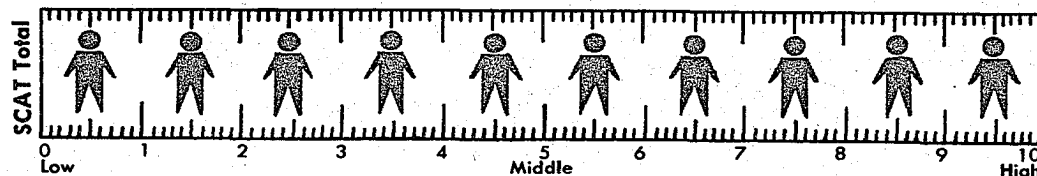
about \_\_\_\_\_ out of 10 students are high



In the kind of ability measured by \_\_\_\_\_

about \_\_\_\_\_ out of 10 students are low

about \_\_\_\_\_ out of 10 students are high



In the kind of ability measured by \_\_\_\_\_

about \_\_\_\_\_ out of 10 students are low

about \_\_\_\_\_ out of 10 students are high

(To find out how to compare a student's own performance on any two tests taken, turn to page 4.)

**STEP**

**Mathematics** . . . measures your ability to understand numbers and ways of working with them (for example, addition and division), such symbols as  $+$ ,  $\sqrt{\quad}$ , and  $<$ , relationships between objects in space, how two changing things can depend on each other (for example, distance and speed), how to draw conclusions from facts, and how to make estimations and predictions when you do not have all the information. Mathematics teachers call these concepts *number and operation*, *symbolism*, *measurement and geometry*, *function and relation*, *deduction and inference*, and *probability and statistics*.

**STEP**

**Science** . . . measures your ability to recognize and state problems relating to science, to select ways of getting information about the problems, to understand and judge the information you get, to predict what the solutions to these problems may be, and to work with symbols and numbers used in science problems. Some of the questions are about biology materials; some are about chemistry, physics, meteorology, astronomy, and geology. All of the questions present science in practical situations (for example, in the home, on the farm, and at work).

**STEP**

**Social Studies** . . . measures your ability to understand the kinds of social studies materials which a citizen in a democracy should be able to deal with. These include maps, graphs, cartoons, editorials, debates, and historical documents. There are questions about history, geography, economics, government, and sociology.

**STEP**

**Reading** . . . measures your ability to read materials and then answer questions about what you have read. These questions ask you to remember specific things the author said, to understand what he meant and why he might have said what he did, and to criticize his ideas. The reading materials include directions, announcements, newspaper and magazine articles, letters, stories, poetry, and plays.

**STEP**

**Listening** . . . measures your ability to listen to materials and then answer questions about what you have heard. The Listening test is very much like the Reading test except, of course, you *hear* instead of *see* the things you are asked to remember, understand, or criticize.

**STEP**

**Writing** . . . measures your ability to criticize materials written by other students in terms of the ways they are organized or written. The questions ask you to pick out errors or weaknesses in the writing and choose revisions which best correct the errors or weaknesses. The materials were written by students in schools and colleges in various parts of the United States; they include letters, answers to test questions, school newspaper articles, announcements, essays, outlines, directions, and stories.

**SCAT**

**Verbal** . . . measures your ability to understand sentences and give the meanings of words. This ability is most important in such school courses as English, foreign languages, and social studies (history, civics, etc.).

**SCAT**

**Quantitative** . . . measures your ability to perform operations with numbers and to solve mathematics problems stated in words. This ability is most important in such school courses as mathematics and science.

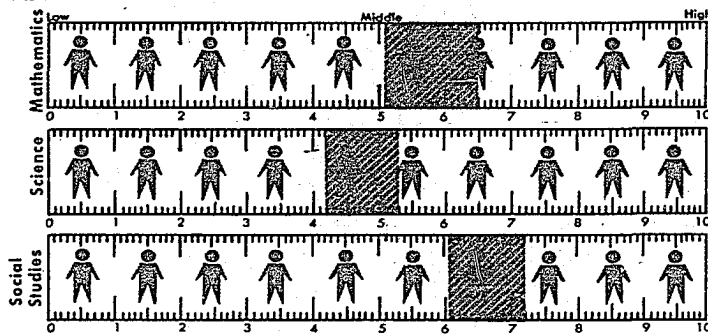
**SCAT**

**Total** . . . combines your scores on SCAT Verbal and SCAT Quantitative to provide the *single* best measure of your general capacity to do the work of the next higher level of schooling.

## How to compare a student's performance on any two tests taken

1. If the shaded areas for any two tests overlap, it is impossible to say with any certainty that the student's standing on one test is higher than his standing on the other test.
2. If the shaded areas for any two tests do *not* overlap, one can say with considerable certainty that standing represented by the area farther to the right is higher than standing represented by the area farther to the left.

### EXAMPLE

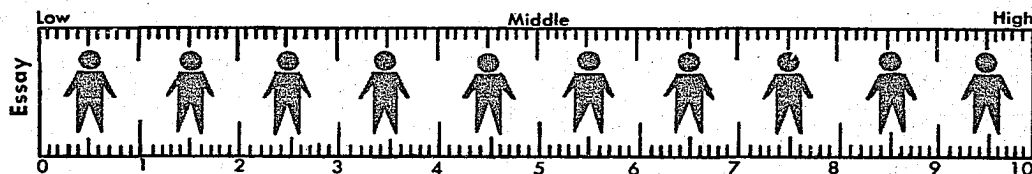


The shaded areas for Mathematics and Social Studies overlap; there is no important difference in standings on these two tests. The same is true of Mathematics and Science. However, the shaded areas for Science and Social Studies do not overlap. The student is higher in Social Studies than in Science ability, as measured by these tests.

Conclusions about comparative performance:

**and STEP Essay**...measures your ability to write a short explanation, letter, account of a personal experience, argument, or discussion of a problem. Your essay is compared with essays on the same topic written by a large group of students at your grade level.

\_\_\_\_\_’s standing on STEP ESSAY is indicated by  
(name)  
the shaded area. Free-writing ability, as measured by this test, is compared with that  
of students in grade \_\_\_\_\_ in the \_\_\_\_\_  
(grade) (publisher's group or local group)



about \_\_\_\_\_ out of 10 students are lower

about \_\_\_\_\_ out of 10 students are higher

STEP/SCAT SUPPLEMENT TO  
STUDENT REPORT FOLDER

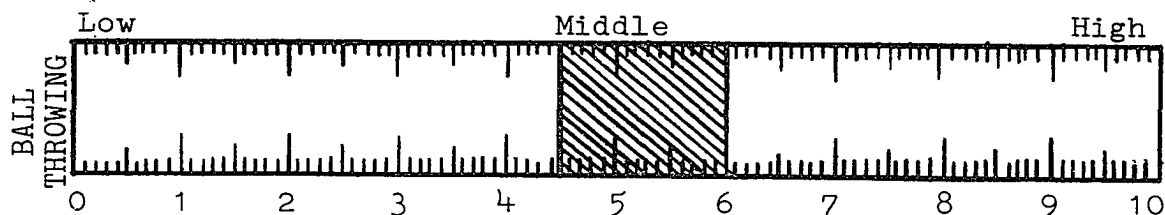
This report has been prepared to help you understand the results of the ability and achievement tests you took recently. Understanding your scholastic achievement and ability will be of value to you in making your future educational and vocational plans. Knowing the meaning of the following terms will help you understand your test results.

- ABILITY: A measure of aptitude or capacity for learning.  
(How much can I learn?)
- ACHIEVEMENT: A measure of mastery of some particular subject area.  
(How much have I learned?)
- PERCENTILE (%ile) RANK: Indicates the percentage of students, with whom you are being compared, who have scores equal to or lower than yours. For example, a percentile rank of 70th percentile shows that 70% of the students had scores the same as yours or lower and that about 30% had scores higher than you did.

You probably have learned that no test, regardless of what it measures, will give you the same results everytime you take it. If you were asked to throw a ball as far as you can 10 times, you would probably throw it a different distance each time. However, it is possible to determine a range (a band of distances) that is characteristic of your skill in ball-throwing. This could then be compared with other students your age to see how you compare with them in this skill. In a like manner, if you had taken each of the achievement and ability tests 10 times, you would probably make a slightly different score each time. For this reason, a range (band) of scores has been determined on each of the tests you took based on the scores you actually made. These scores will be reported to you in "percentile bands". These bands characterize your performance in several achievement and ability areas as compared to a nation wide sample of students in your grade.

For an example, let us return to the skill of ball-throwing. Suppose that when you were tested on this skill you made a throw of 62'. From previous experiments it is possible to determine that if additional throws were made the majority of them would range from 56' to 68'. When this is compared with the performance of other students your age it is found that your band (56' to 68') gives you a percentile band of 45-60. This places you higher than 45 percent of the

students and lower than 40 percent (the difference between the upper limit of the band and 100). A percentile band of 45-60 would appear on a scale as follows:



When is a Difference an Important Difference?

It is also possible to determine when a difference between two scores is important enough to indicate a true strength or weakness. The example on page 4 (the back) of the Student Report folder explains how these differences may be determined. You should pay special attention to any important difference between (1) your QUANTITATIVE ability (SCAT Quant.) and your achievement in MATHEMATICS, (2) your OVERALL ability (SCAT Total) and your achievement in SCIENCE, SOCIAL STUDIES, and WRITING and, (3) your VERBAL ability (SCAT Verbal) and your achievement in READING. These comparisons will help you see in which subject areas you might be expected to show the most improvement.

DO NOT OPEN YOUR STUDENT REPORT FOLDER UNTIL TOLD TO DO SO



APPENDIX C

TYPESCRIPTS

TYPE-SCRIPT: REPORTING STEP/SCAT TEST  
RESULTS TO STUDENTS IN GROUPS

COUNSELOR: We are ready this morning to have a look at the results of the tests which you took at the beginning of last month. I'm sure that you all will be interested in them, but before we actually have a look at the results, I want us to think just a moment about what measurement means to us in all areas of life. I'm sure that if we stop to think, we'd be surprised at how often we use measurement as a reference in our daily thinking and conversation. We have, of course, the measurement of time. You're very, very conscious of that in school at least. We may have been saying just a few days ago, "Oh, just three more days until the Homecoming game," or you might be thinking each day or each period, "Ten more minutes until the bell rings." I think each of us is probably thinking about now, "Just one more week until the Thanksgiving holidays." In addition to time as a measurement, of course we also have another form of time, and that is age. Age means a number of things to you. For instance, when you were about six years of age, somewhere in the vicinity of that, you started to school. And that made quite a difference in your life, didn't it? You couldn't goof off around the house anymore, play in the sand box; you had to get with it, get to school, get something done. So this made a difference in your routine. Another measure that we use quite a lot is the measure of distance. It's so far from here to another place.

That may have a direct bearing on whether you go or not. One way we might be able to see that distance has a direct bearing upon what we do--let's assume that Suzy over here lives about three miles from school, and she rides the bus while Joe lives about a block and a half from school. Now then, Suzy is going to have to get up, have breakfast, find her books, get everything together, and probably catch the bus at seven-thirty in order to get here to school on time. Maybe Joe will be able to sleep until seven-thirty, jump up and grab a bite, and still be over here in time for school. This, as a measurement then, has a direct effect upon the way these people live their lives, a direct bearing upon them. Some other measurements are physical measurements. You are a mass of physical measures. You have a head size. You boys especially have a neck size and an arm size for your shirts; girls have dress sizes, and of course, your hose sizes and shoe sizes, glove sizes. All of these different measurements go to make up you. They are points of reference to really bring us a picture, a better picture, of you. And the more accurately you know your own measurements the better you are able to shop for your apparel, aren't you? Another kind of measurement can help us in our planning and shopping in a different way. And that is a measurement of your academic achievement and a measure of your academic capabilities. In this way you can make more realistic and better plans for your future, for your occupations, for the courses you may take

next year, and the year following. So this is really what we are trying to get at when we give you measures from various kinds of tests. It's to help you to understand yourself a bit better. Now then, if you'll look here on the very first page of your Student Report Folder, you'll see that you are given the names of the tests that you took. Of course, there were quite a number of them. Some of the tests were called the Sequential Tests of Educational Progress, abbreviated to STEP. These were the achievement tests. They are designed to help you keep track of your own individual development in school and college. They will also help your teachers guide you in learning the important skills and processes that you go to school to learn. There are seven kinds of STEP tests: Math, Science, Social Studies, Reading, and Listening (which we did not take), Writing, and Essay Writing, and of course, we did not take the Essay Writing either. Some of the facts which you would like to know about STEP are that STEP tests measure how well you are able to use what you have learned. In other words, if you actually understand something that you have learned, then you should be able to use this knowledge, this learning, towards the solving of new problems. This is what we're trying to measure in tests such as these achievement tests. Another important fact that I'd like for you to remember is that the skills which the STEP achievement tests measure are skills which can be improved if you work at improving them. Now then, the second, no,

actually the first test that we took but the second listed here on the page is the SCAT, named the School and College Ability Tests. There were two of these. One was designed to measure your ability to use and understand words. This we term Verbal ability. The second was designed to measure your ability to use and understand numbers, and this we call Quantitative ability. Now, I'd like for you to remember that term because we are going to refer to it again, and it may be new to you. The ability to use and understand numbers is referred to as Quantitative ability. Why just these two abilities, you say? "I'm not very good here, but I have a lot of mechanical ability, or I have a lot of artistic ability," or something of the sort. That's good. These things may mean a great deal to you in your life, but as far as your class work is concerned and your actual school courses are concerned, these two abilities, Verbal and Quantitative, have the greatest bearing, and so these are called the School and College Abilities Tests. Now, of course, there are STEP and SCAT tests for appropriate grade levels. Those of you who have been in the \_\_\_\_\_ City system will recall that you took these tests, at least some of them, at the seventh and eighth grade levels. And, of course, this is another reason for a continuation of the testing program. That is in order that you may follow your individual development. Do the same strengths and/or weaknesses show up again and again, year after year? Both the STEP and the SCAT tests were designed to see how

well you can work, not how fast. They were not speed tests, and I'm sure that most of you had sufficient time to complete the tests if you attempted to do so. Now then, without opening the Report Folder, slip the inside page out. This is a supplement which has been prepared to help you understand more clearly the results of your tests. "Understanding (this is the second sentence of the paragraph if you would like to read it)--Understanding your scholastic achievement and ability will be of value to you in making your future educational and vocational plans." Now, knowing the meaning of the following terms will help you to understand your test results. First, ABILITY. What do we mean by ability? We see here that it is defined as a measure of aptitude or capacity for learning. We might sum it up in the question, "How much can I learn?" Now then, the second term, ACHIEVEMENT, is defined here as a measure of mastery in some particular subject area. We might sum this up as, "How much have I learned?" Now, you've been going to school, most of you, for a little over nine years. That's quite a bit of school. How much have you learned in this length of time about mathematics or about science; or, how much have you learned about using and understanding words in various ways? This is what we are trying to find out when we are measuring achievement. Now, the third term that we need to understand is PERCENTILE RANK. This indicates the percentage of students with whom you are being compared who have scores equal to or lower than yours.

For example, a percentile rank of the 70th percentile shows that 70 percent of the students in the group with which you're being compared would have scores the same as or below yours, and only about 30 percent would have scores higher. Now then, you've probably learned that no test, regardless of what it measures, will give you the same results every time. If you were asked to throw a ball as far as you can ten times, you would probably throw it a different distance each time. However, it is possible to determine a range or a band of distances that is characteristic of your skill in ball throwing. This could then be compared to other students of your age in the kinds of tests you have here, in your grade level, to see how you compare with them in this skill. In a like manner, if you had taken each of the achievement and ability tests ten times, you would probably have made a slightly different score each time. For this reason, a range or band of scores has been determined on each of the tests you took, based on the scores that you actually made. These scores will be reported to you in Percentile Bands. These bands characterize your performance in several achievement and ability areas as compared to a nationwide sample of students at your grade level. So you see, your comparison here is being made, not with just tenth graders in \_\_\_\_\_ City, nor with tenth graders in Oklahoma City or the State of Oklahoma, but actually is based on a sample (and by sample that doesn't mean two or three but hundreds and hundreds) of tenth grade

students all across the country. How do you stack up? How do you rank? Where in this group would you line up, you see. What position in this group would you have on the basis of these tests you took? It's a comparative thing. We cannot equate a 70th percentile with a grade of seventy, nor a 50th percentile with a grade of fifty. It does not mean the same thing. But where in the group, you see, do you stand with regard to the score that you made? We don't know the exact score; it doesn't really matter, because we're getting a comparison here within a group. I do want you to understand that it is not the same thing as you are accustomed to thinking of as a grade. But remember, the rank shows what proportion. If it's a 60th percentile then that means that you are here, and 60 percent of the group with whom you are being compared would have the same score or a score lower than yours. And 40 percent would have scores higher. If it were the other way around and you are at the 40th percentile, then that means that 40 percent of the group would have scores equal to yours or lower and 60 percent (approximately) would have scores higher. Are there any questions about this now? Okay, fine. Now then, with reference to the bands reported here--rather a percentile point. You remember last year on your Differential Aptitude Tests you had a particular percentile rank or point in Verbal Reasoning and Numerical Ability and Mechanical Reasoning and so forth, and the little graphs came out with lines connecting. This is a band as



explained in the paragraph we read. Let's read this paragraph just above the example here, and I think that will explain it for you. "For an example, let us return to the skill of ball throwing. Suppose that when you were tested on this skill you made a throw of 62 feet. From previous experience it is possible to determine that if additional throws were made, the majority of them would range from 56 feet to 68 feet. When this is compared with the performance of other students your age it is found that your band (56 feet to 68 feet) gives you a percentile band of 45 to 60, meaning that you are higher than 45 percent of the students in ball throwing and lower than 40." Yes, lower than 40, you see? Okay, now look at the little band here, because this is the way it is going to be shown. Notice here it'll be on a scale like this. This shaded area would be a percentile band of 45 to 60. Any question about that? You can see now how it is read. Now then, when is a difference an important difference? "It's also possible to determine when a difference between two scores is important enough to indicate a true strength or weakness. The example on page four of the Student Report Folder explains how these differences may be determined." All right now, flip right over to the very back of your Student Report Folder and here at the top we'll see an area which says, "How to compare a student's performance on any two tests taken." First, if the shaded areas for any two tests overlap, it is impossible to say with any certainty that

the student's standing on one test is higher than his standing on the other test. Now, look down at the example. You see you have a scale first for mathematics, and let's notice the percentile band there, 51 to 65, notice the second scale for science, 42 to 53. Do we have an overlap there? Yes, we do, a slight one; but there is an overlap. Referring again to point one, if the shaded areas of any two tests overlap, it is impossible to say with any certainty that the student's standing in one area is higher than his standing in the other. Now, according to the shaded areas, it would appear that his standing is higher in mathematics, right? But according to our instructions here, we cannot really say that this student's standing would be any stronger or any higher in math than in science, because this overlap accounts for that. It indicates no significant difference here, you see, between these two tests. All right, now let's read number two. "If the shaded area for any two tests do not overlap, one can say with considerable certainty that standing represented by the area farther to the right is higher than the standing represented by the area farther to the left."

Notice your social science scale here with the percentile band of 61 to 72. Does that overlap then with science? No, not at all. We have a significant difference here. So we could say with confidence then that this student's ranking in the area of social studies was higher than his standing in science. Now this is what it means and now to look on your

own tests to see your particular strengths or weaknesses. These comparisons will help you to see which subject areas you might be expected to show more improvement. Let's open the page; open it all the way, because I want you to read across. Notice here now, you have a number of bands. The first six are the STEP tests; of course, we took only five. We did not take the Listening which is listed here, but these are the STEP tests. These are the achievement tests which attempt to answer, "How much have I learned in these various subject areas?" And the lower three are those measures from the SCAT or the abilities tests. How much can I learn within these two areas, and then a total which we are going to talk about. Now then, look right across the page from each scale, and let's go over very quickly what kind of thing we've attempted to test here. In mathematics--notice right across here--STEP Mathematics measures your ability to understand numbers and ways of working with them. For example, addition and division, various mathematical symbols, relationships between objects and space, how two changing things can depend on each other, for example, distance and speed, how to draw conclusions from facts, how to make estimations and predictions when you do not have all the information. These are the concepts which math teachers call number and operation, symbolism, measurement in geometry, function in relation, deduction and inference, and probability and statistics. So you see, they took a rather deep measurement here, didn't they?

The Science test measures your ability to recognize and state problems relative to science, to select ways of getting information about the problems, to understand and judge the information you get, to predict what the solutions to these problems might be, and to work with symbols and numbers used in science problems. Some of the questions are about biology materials, some are about chemistry, physics, meteorology, astronomy, and geology. Did you know you knew anything about those? All of the questions present science in a practical situation, for example, in the home, on the farm, and at work. So you do know something about these various areas. You haven't taken advanced study in many of them, of course, but you do know many fundamentals, because that has been incorporated in your learning, you see, as you've come along through school. And if you'll think about it for a minute, I know that you'll remember that it has. STEP Social Studies measures your ability to understand the kinds of social studies materials which a citizen in a democracy should be able to deal with. These include maps, graphs, cartoons, editorials, debates, and historical documents. There are questions about history, geography, economics, government, and sociology. Now then, the Reading test measures your ability to read materials and then answer questions about what you have read. These questions ask you to remember specific things the author said, to understand what he meant, why he might have said what he did, and to criticize his ideas. The reading

materials include directions, announcements, newspaper and magazine articles, letters, stories, poetry and plays. We did not take the STEP Listening. The Writing test measures your ability to criticize materials written by other students in terms of the way they are organized or written. The questions ask you to pick out errors or weaknesses in the writing and to choose revisions which best correct the errors or weaknesses. The materials were written by students in schools in various parts of the United States. They include letters, answers to test questions, school newspaper articles, announcements, essays, outlines, directions and stories. And I'm sure you remember a number of these things. Now then, we get to the SCAT. Let's see what we have here in the measurement of abilities, school and college abilities. Remember we had the two areas, Verbal, which measures your ability to understand sentences and give meanings of words. This ability is most important in such school courses as English, foreign languages, and social studies, such as civics and history and so forth. Remember now that these abilities have definite bearings upon course selection whenever possible, and this is the kind of thing we need to look for--each of you in your own particular measurements. All right, now then. SCAT Quantitative means what? Ability to deal with numbers. All right, good. Your ability to perform operations with numbers and to solve mathematics problems stated in words. This ability is most important in such school courses as

mathematics and science. Now then, the SCAT Total down here combines your scores on SCAT Verbal and SCAT Quantitative to provide the single best measure of your general ability to do the work of the next higher level of schooling. So this is what it's all about. Let's look at your scores, and as I call out these various things, I want you to compare. Incidentally, this is listed at the bottom of your supplement sheet so you can do it yourself later on. Right now I want you to compare your Quantitative measure, that is your SCAT Quantitative which is the second scale from the bottom of the page with your Mathematics achievement test which is the very top scale. What do you see? Do you have an overlap there between your percentile bands? If you do, of course, then this indicates that you have--you show a consistency here between how much can I learn and how much have I learned. Right? Now then, do you find that they do not overlap? If so, then the area farthest to the right--higher at the scale--whichever scale that is--shows that you measure on these tests to have a higher standing than you did on the one to the left. If they do not overlap, this is to be appraised as a significant difference. All right, now let's check your SCAT Total which is the bottom one and the best single measure of your School and College Ability. Check that against your science achievement which is the number two scale from the top of the page. Is there an overlap? If there is an overlap, remember what that means. Referring to the back of the page, we

cannot say that your standing in any one area is higher than the other, but if it does not overlap, then you do have, at least as measured by this test, a significant difference here. Your standing in the area which has the percentile band farther to the right, farther upscale, you see--then you are higher in that area than you are in that one to the lower end. Do you understand? Do you have a question? (No questions.) Now let's compare your SCAT Total again with Social Studies. What have you here? Notice your percentile bands; notice the lower limit and the upper limit of each percentile band. Is it from the 40th to the 60th? Is it from the 50th to the 70th? Or, is it from the 20th to the 40th? Wherever it is, notice the number and then you can tell more readily whether you have an overlap or not. All right. Measure your SCAT Total against your Writing which is the last one of the achievement tests, number four from the bottom. You may possibly find this to show the greatest difference, but I don't know; you can determine that. This is your measurement, and it's for your consideration. Now we have one more area that we need to compare, and that is the SCAT Verbal which is number three up from the bottom of the page and your Reading band. Now, are there any questions regarding the meaning of any of your scores? If so, raise your hand, and I will come to you and see if I can help you understand them.

After a brief period of questions, the test profiles were collected, the criterion instrument was marked, and the group was dismissed.

TYPE-SCRIPT: STEP/SCAT RESULTS  
TO THE INDIVIDUALS

C1--Hello, how are you today?

S1--Oh, okay, I guess.

C2--Are you enjoying this weather we're having?

S2--Well, it's kinda cold, but I guess it's okay.

C3--The reason I called you in today is because I have received the results from the tests that we took last month.

S3--Oh, good grief!

C4--Before we actually look at your scores I'd like for us to review just a moment the kinds of tests we took. As you see there on the front page, (points to cover page of the Student Report Folder) the names of some of these tests--

S4--(interrupts) Yeah--

C5--are the Sequential Tests of Educational Progress which we abbreviate and call STEP. Now these are achievement tests which are designed to help you keep track of your own individual development in school. They also help your teachers to guide you in learning the important skills and processes which you go to school to learn. As you may recall we took five of these tests: Mathematics, Science, Social Studies, Reading and Writing. We did not take Essay Writing and



Listening. Now some of the facts I'd like for you to remember about STEP are: the tests are designed to measure how well you are able to use what you have learned about a subject; then you should be able to use that learning in solving new problems in that field. The skills measured by STEP tests are skills that can be improved if you work at improving them, and this is an important thing to remember. Now uh another kind of test that we took, in fact it was the first one that we took, you may remember, the SCAT or the full name, The School and College Ability Test. And these were designed to help you and your teachers estimate your general ability to do the work of your next school year. They do this by measuring your ability to use and understand words; this we call Verbal ability and by measuring your ability to use and understand numbers. This is called quantitative ability. Now I want you to remember the term because we are going to use it again. The reason that these two abilities are measured is because these are the abilities that you need most in order to succeed in school. It might be that you have a great deal of artistic talent, let us say-- S5--(interrupts) Huh (laughs)-- C6--ability. Or perhaps you have a lot of mechanical ability. These are very good things to have but they don't necessarily help you a great deal in doing school work, as you well know. The greatest portion of your classroom work depends on your abilities to use and understand words and numbers. Now then,

both STEP and SCAT tests are uh, come in various appropriate grade levels.

S6--Uh-huh.

C7--You probably remember--

S7--(interrupts) Yes--

C8--when you were in the seventh and eighth grades--

S8--(interrupts) Umm--

C9--that you took STEP--

S9--(interrupts) It took all day--

C10--and SCAT tests over there. The uh--both these tests, the STEP and the SCAT are not speed tests. They are not designed to see how fast you work, but rather to see how well you can work. These tests are part of carefully planned educational programs. The results are for your own benefit. They'll help you to see your strengths and weaknesses and once you know these, you can decide more wisely upon what subjects you should work harder and what courses of study you might decide to take.

S10--Well, the uh teachers later on--I mean like next year and so on, they see these, don't they?

C11--Well, possibly, they might refer to them, but I think it's more important that you keep your own Student Report Folders from year to year. Remember last year you received a profile from your Differential Aptitude Tests.

S11--Oh yeah. (laughs)

C12--And each year as you go through your school career if

you'll keep your test results, your scores, then I think that you will be able to make comparisons. Now then, without opening your folder just slip this supplement sheet right out of it, if you will--

S-12--Is this it?

C13--Yes. Now this has been designed to help you understand your test scores a little bit better, and of course if you can understand your scholastic achievement and ability this will be of great value to you in making your future educational and vocational plans. Knowing the meaning of the following terms will help you to understand these results. First, ability--what do we mean by ability? Here we see it defined as "a measure of aptitude or capacity for learning. In other words, we might sum it up in the question, how much can I learn? Achievement on the other hand, is a measure of mastery of some particular subject area. We might sum that up in the question, how much have I learned? Now you've been in school for approximately nine plus years, I imagine, more or less. At the very beginning you started in to learn how to use words and how to use numbers. Well now, in these nine years, how much have you learned? How far along the scale have you come? This now is what the tests are attempting to measure. Now the third thing that we need to know is what we mean by percentile rank. This indicates the percentage of students, with whom you are being compared, who have scores equal to or lower than yours. For example, a percentile rank of 70%ile

shows that 70 percent of the students had scores the same as yours or lower and that about 30 percent had scores higher. Do you understand now?

S13--Uh-huh. Yeah.

C14--Now then uh, as you know, you probably would not make exactly the same scores on these tests if you took them over again. If for example you took all these tests as many as ten times, you might very well come up with ten different scores on each one. A little comparison is made here which you can read later on at your leisure--about ball throwing.

S14--Um-hum. They had that in our test, that same kind of deal, in other words, they were explaining this percentile and they gave us the same kind of example.

C15--Oh yes. I believe that was on the questionnaire, was it not?

S15--Yeah. That was on the SCAT..

C16--Oh, was it?

S16--Before we started SCAT, they--

C17--(interrupts) Oh, I see. Well, at any rate, in order to review, if you were to throw a ball ten times it would probably land in ten different places. (pause)

S17--Um-hum.

C18--But after a while you would be able to say, or decide, that the ball would be most likely to land between here and here, let us say. So along this same line you'll find that your scores are given this time in percentile bands rather

than at a single percentile point. Notice here on the little example: We find a percentile band from 45 to 60. So this would be the (pause) shaded area--would be considered your range--

S18--(interrupts) Uh-huh--

C19--of scores, somewhere in between that, between the lower and upper extremities.

S19--Well, how, how can I figure that out if I've only taken the test once? In other words, is that the way it's going to be marked off on our, in our little sheet here?

C20--Yes, it will be. We'll see that in just a few moments. Well, I don't know exactly the technicalities, but it has to do with how consistently the test measures and--

S20--(interrupts) Um-hum--

C21--so forth. All right, now then it is important for us to remember one other thing. Notice right under the example: when is a difference an important difference? It is also possible to determine when a difference between the two scores is important enough to indicate a strength or weakness. The example on page 4 of the Student Report Folder explains how these differences may be determined. Now then, turn your Report Folder right over on the back and we'll see here at the very top how to compare a student's performance on any two tests taken. The first instruction: "If the shaded areas for any two tests overlap it is impossible to say with any certainty that the student's standing on one test is higher

than his standing on the other." Now then, let's look at the little example here. You notice a scale for Mathematics, with a percentile band from 51 to 65.

S21--Um-hum.

C22--Now then, notice the second on Science. And what is the percentile band there? From what?

S22--It starts 42 to, it looks like 53.

C23--All right.

S23--It overlaps there.

C24--All right, there is a small overlap. So what must we remember? (pause) It is impossible to say with any certainty that this student's standing in Mathematics is any higher than his standing in Science, even though it might appear so by--

S24--(interrupts) Um-hum--

C25--by the band. But since they overlap the difference is not significant enough to be important, you see. All right, now then the second instruction: "If the shaded areas for any two tests do not overlap one can say with considerable certainty that the student's standing represented by the area farther to the right or up the scale," you see? (points) "is higher than the standing represented by the area farther to the left." Now let's drop down to the third scale, here. Social Studies. Notice here the percentile band from 61 to 72. Now compare that with the score or the percentile band in Science. Is there an overlap?

S25--No.

C26--No. There is no overlap. So in this case we would be able to say with considerable certainty that this student's ranking in Social Studies was higher than his ranking in Science.

S26--Um-hum.

C27--But not higher necessarily than his ranking in Mathematics, you see, because those two do overlap. Okay, now, do you think you understand this?

S27--Um-hum.

C28--All right, let's look inside at your scores. Your scores compare you to a nationwide sample of tenth grade students. Just open it right up. And there you see the shaded areas. Now first, before we begin to talk about that too much, I'd like for you to notice--right across from each band--that there is an explanation of just what measurement was attempted in each area. These first six scales, those with the subject areas listed; that the left side are the scores from the STEP tests, the achievement tests. These, in other words, attempt to answer how much have I learned, how far have I come, you see, in these various subject areas? And these last three represent scores in the SCAT or the School and College Ability Tests. All right, now then uh--let's notice here on the opposite page, in Mathematics, the measurement of your ability to understand numbers and ways of working with them, for example, addition, division and so

forth, and various Mathematical symbols, relationships between objects in space, how two changing things can depend upon each other, how to draw conclusions from facts, and how to make estimations and predictions when you do not have all the information.

S28--Um-hum.

C29--Now then, in Science they attempted to measure your ability to recognize and state problems related to Science, to select ways of getting information about the problems, to understand and judge the information you get, to predict what the solutions to these problems may be and to work with symbols and numbers used in Science problems. Some of the questions are about biology materials, some are about chemistry, physics, meteorology, astronomy and geology. Did you know you knew something about all of those?

S29--No. (laughs)

C30--All right now I think this next sentence explains that. "All of the questions present Science in practical situations. For example, in the home, on the farm and at work." So if you'll think about it in the light I'm sure you will--

S30--(interrupts) Yeah I remember--

C31--realize that you do know something about these various fields. All right, now in Social Studies the measure was of your ability to understand the kinds of Social Studies materials which a citizen in a democracy should be able to deal with. This included maps, graphs, cartoons, editorials,



debates and historical documents. There were questions about history, geography, economics, government and sociology.

"Reading measures your ability to read materials and then answer questions about what you have read. These questions ask you to remember specific things the author said, to understand what he meant, and to understand why he might have said what he did and to criticize his ideas. The reading materials include directions, announcements, newspaper and magazine articles, letters, stories, poetry and plays." Now we did not take the STEP Listening test, as you probably remember-- S31--(laughs) That's okay.

C32--And in Writing, the measure of your ability was in the area of your ability to criticize materials written by other students in terms of the ways they are organized or written. The questions asked you to pick out errors or weaknesses in the writing, to choose revisions which best correct the errors or weaknesses. The materials were written by students in schools and colleges and so on and so forth. And then this uh will remind you of various things that were included in the questions. Now then to go to the SCAT. Notice SCAT Verbal; remember we had two measures in the School and College Abilities; SCAT Verbal measures your ability to understand sentences and give the meanings of words. This ability is most important in such school courses as English, foreign languages and Social Studies.

S32--I think this is the only time that Latin paid off.

C33--Do you?

S33--Because I remember two or three words in there that I just didn't know the meaning and I knew that I could tell that they looked like a Latin word that we'd had and so I just kinda (laughs) guessed, and I guess I got pretty good.

C34--All right, now then. STEP Quantitative measures your ability to perform operations with what?

S34--Numbers.

C35--Numbers. All right, let's remember that Quantitative refers to numerical ability, and to solve Math problems stated in words. This ability, of course, is most important in such school courses as Math and Science. Now then, the SCAT Total down here combines your scores on SCAT Verbal and SCAT Quantitative to provide the single best measure of your general capacity to do the work of the next higher level of school. Referring again to the supplement page--you'll notice here in this last paragraph, the statement, "You should pay special attention to any important differences between--" now if you want to refer to your scales as I call them off, let's uh--

S35--(interrupts) At the bottom here?--

C36--Um-hum. All right. Your Quantitative ability, now where is that?

S36--The middle one.

C37--SCAT, all right. And your achievement in Mathematics?

S37--Up here?

C38--Yes, the very top one. Now then, do you have an overlap

there?

S38--No.

C39--Do the percentile bands overlap?

S39--No. Huh-uh.

C40--All right, which is the higher?

S40--Mathematics is the higher one.

C41--What is the lower limit in your percentile band in the Math Achievement? 80? I can't see from here.

S41--Yes, 80.

C42--80. And the upper limit in the SCAT Quantitative?

S42--75.

C43--75. So then you do not have an overlap there. You show a higher score then in the achievement test, right?

S43--Um-hum.

C44--In your achievement in Mathematics.

S44--Were these right here what we took--that little short page or that first one?

C45--The SCAT, you mean?

S45--Yes.

C46--Well, we did take the SCAT first. It was all in one booklet, I believe.

S46--And they graded those, too, in other words, by what you checked? Now is that the same thing as--

C47--(interrupts) Now you're not mistaking this for the questionnaire? The very first thing we did was the questionnaire, where you just marked boxes and so on and so forth.

Now that was not a test. But then we had the SCAT test and I believe it was all in one booklet, but part of it was Verbal and part of it was Quantitative.

S47--These are STEP, right? (points to upper part of SRF)

C48--Yes, achievement tests, remember? They were separate booklets.

S48--Um-hum.

C49--Pertaining to subject matter areas.

S49--Oh, I see.

C50--Okay.

S50--It's been so long ago. (laughs)

C51--So then you would show greater strength of achievement than do most tenth graders with reference to their measured ability to achieve in the Quantitative area, you see?

S51--Well, this is my measurement, is that right?

C52--Of your ability.

S52--Yes.

C53--On this test, yes.

S53--That was on the SCAT. And this is also what I made.

Now--

C54--(interrupts). On the achievement test in Math. This was the separate booklet; we rotated the booklets.

S54--Yes.

C55--Okay. These were the STEP tests, then. Up here.

S55--Yeah.

C56--And this was the very first test we took.

S56--I remember that one, yeah.

C57--Okay. All right, let's go on to the next one. Perhaps we can come back to that one after we look at the others.

S57--Okay.

C58--Now then, I want you to compare your overall ability, SCAT Total, with your Science achievement.

S58--Um-hum.

C59--Is there an overlap?

S59--Yes.

C60--All right, then we cannot say that you are any higher or stronger in the one area than in the other, right?

S60--That is, on the SCAT Total, is it they averaged my two?

C61--I don't think it's exactly an average, but it is a combination. At any rate, because an average would not come out exactly that way if it were just a straight Mathematical average. But it is a combination. Notice here, combination of your scores, in order to provide the single best measure of your capacity. So then we find a consistent performance here you see between your achievement and your measured ability to achieve. All right now, compare your SCAT Total with your Social Studies.

S61--(laughs).

C62--Do you have an overlap?

S62--Huh-uh. No!

C63--Which is the higher?

S63--My total!

C64--Your SCAT Total, your abilities total, then, shows higher

with no overlap with your Social Studies. Now what do you understand from that?

S64--Well, uh, (laughs) well, evidently I'm not (pause) well, I don't really know, I can't quite, (pause)

C65--Remember, these tend to answer how much have I learned, and these how much can I learn. You see? So (pause)

S65--Oh.

C66--You appear to be able to learn more than you actually have learned in your Social Studies, see? In other words, this indicates that there is possible room for improvement in this area.

S66--Well, would there be anything (pause) the last time I had Social Studies was in seventh grade, and uh (pause) would that have anything (pause) I kinda figured it probably had something to do with that (pause) if I have more ability to learn, which it shows down here, and you know (pause) I have up here, I guess about three years ago, and so (pause, laughs) I, you know (pause)

C67--Haven't you had Oklahoma history yet?

S67--Well, yes, (pause) but I mean uh (pause) well (laughs, pause) none of the questions really concerned Oklahoma history. It was more over World history and I--

C68--(interrupts) I understand. So you feel that it's more a situation that you haven't actually studied it lately?

S68--Well, yeah (pause) I haven't had a chance to take it. (laughs)

C69--Well, that may be true. Now let's go on to something else. Now measure your SCAT Total with your Writing band. Do you have an overlap?

S69--Yes.

C70--All right. Then here we cannot say that you are higher in one test than in the other if you have an overlap. So this would tend to show a consistent performance.

S70--On these you really can't tell exactly what you did. I mean uh (pause) how well you do. You've got to just kinda guess it's between this. (Points to a band)

C71--In between? Well, that's true, but let's remember what a percentile is. With the percentile band here, let's take for instance your Writing test that we talked about-- (pause).

S71--Okay.

C72--This would indicate that you were--that you stood with or above?

S72--58 percent.

C73--58 percent of the group with whom you are being compared and below?

S73--85.

C74--Well, no not 85. You see, you have to start from this end, so that would be fifteen percent. See, because your top measurement is 85.

S74--Oh, yeah, yes.

C75--See? Do you understand now?

S75--Um-hum.

C76--Okay. These little (pause) these little men are set up as if you were compared to ten, see.

S76--Um-hum.

C77--If you were one out of ten, where would you be? So this would mean that you were six--six out of ten would make scores lower than you and one would make scores higher than you in this particular area.

S77--Um-hum.

C78--Now we have one other comparison to make and that is your SCAT Verbal, your ability to use and understand words and sentences and so forth. Compare that with your Reading score, achievement score. That's it. Do you have an overlap?

S78--No.

C79--No overlap. Which is the higher?

S79--My SCAT Verbal. That would be my ability to learn.

C80--So what do you deduce from that?

S80--Well, that I uh (pause) this is what I have learned. Is that right?

C81--Um-hum.

S81--And this is what I can learn. So evidently I uh (pause, laughs).

C82--You can improve somewhat in that.

S82--Yeah.

C83--All right. The (pause) now does this help you to understand anymore the first situation we had with the--



S83--(interrupts) Um-hum.

C84--SCAT Quantitative and the Mathematics? In that case, you appeared to have actually achieved more than your measured abilities would indicate that we might expect. As compared--remember now that this is always a comparison, not to be in any way equated with a grade--a percentile of 70 is not the same as making 70 on the test, you see.

S84--Uh-hum.

C85--But only how you rank in the group, as compared with the group. All right, do you have any questions you would like to ask? (pause) Suppose for just comparison again, we have a look at your Differential Aptitude Scores from last year.

S85--(groan) Ohhhh.

C86--Now remember, what is ability? Ability is a measure of aptitude or capacity for learning. Now this was a different kind of test designed to measure several different kinds of aptitudes, among them were Verbal Reasoning which should correspond with what they attempted to measure with SCAT Verbal. How does this stack up? How does it compare with your SCAT results?

S86--Well, uh--

C87--Percentile band there in the SCAT Verbal.

S87--Yes. Well, my lowest band is 84. In other words, I'm above 84 percent and I'm below 7 percent.

C88--All right. Your band is actually from 84 to 90--what?

S88--Three.

C89--93. All right, notice last year on your differential aptitude you had a Verbal Reasoning percentile rank of 90.

S89--It's about the same.

C90--You see that would fall within this particular band.

S90--Um-hum.

C91--So this seems to be a rather consistent measurement.

Now then, the other measurement I think we can compare rather well is the Numerical Ability from your Differential Aptitude Test. What is your band this year on your SCAT Quantitative?

S91--Well, uh (pause)

C92--From what? From what to what?

S92--Well, uh (pause) the band is from 49 to 75.

C93--You notice last year you had a percentile rank of 85, which if we'll compare here with your achievement test, you see would have fallen within the band. You see it might be, for some reason, that SCAT Quantitative this year did not reflect too strongly what you can do.

S93--Uh-hum.

C94--A lot of things depend upon whether you actually understand the instructions, or whether you feel well, or (pause)

S94--Yeah, after you take so many of those tests, you get so tired, you know.

C95--Um-hum. Now here is one more. This is a total. Your Verbal Reasoning and your Numerical ability, and we might see if that corresponds with your SCAT Total.

S95--Right here?

C96--You see, you have a total last year of 90. What is your upper measurement here?

S96--88.

C97--So it's very near, and of course, the thing that might have pulled that down this time could be your SCAT Quantitative.

S97--It probably is.

C98--But this is another reason for keeping your Report Folders and comparing from year to year because just as we said earlier, we know that you wouldn't make the same score on two tests of the length of these tests every time, but I think that you might find, if we had enough tests to use that certain consistencies will show, certain strengths and perhaps certain weaknesses. It would be interesting to follow up now on other tests and see about this Social Studies that you were referring to.

S98--Now this is what I have learned, right?

C99--Right. This is what I have learned (pause) what I can learn.

S99--Okay.

C100--Okay, think you understand?

S100--Um-hum. Thank you.

C101--Okay. Goodbye.

S101--Goodbye.