1992

Assessing the academic achievement of high school special education students

Janice Rumpca Howard
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Howard, Janice Rumpca, "Assessing the academic achievement of high school special education students" (1992). *Master's Theses*. 329.

DOI: <https://doi.org/10.31979/etd.gb7d-g66k>
https://scholarworks.sjsu.edu/etd_theses/329

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 1348690

**Assessing the academic achievement of high school special
education students**

Howard, Janice Rumpca, M.A.

San Jose State University, 1992

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

**ASSESSING THE ACADEMIC ACHIEVEMENT
OF HIGH SCHOOL SPECIAL EDUCATION STUDENTS**

A Thesis

Presented to

**The Faculty of the Division of Special Education
and Rehabilitative Services
San Jose State University**

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

By

Janice Rumpca Howard

May, 1992

APPROVED FOR THE DIVISION OF SPECIAL
EDUCATION AND REHABILITATIVE SERVICES

Kathleen Carmona, Ed.D.

Gilbert Guerin, Ph.D.

Peggy Ann Reed, MA

APPROVED FOR THE UNIVERSITY

ABSTRACT

ASSESSING THE ACADEMIC ACHIEVEMENT
OF HIGH SCHOOL
SPECIAL EDUCATION STUDENTS

by Janice Rumpca Howard

This thesis addressed the question of whether or not special education students put forth their full potential when taking standardized achievement tests. Students were tested with the same mathematical problems in two different test settings: 1) a classroom test situation and 2) a typical achievement test situation. The first testing session took place in a classroom test setting with a teacher made copy of the Stanford Diagnostic Mathematics Test (SDMT) and the second session was in a typical achievement test setting with the official SDMT test booklets.

Approximately 25% of the students tested scored higher than 1 standard error of measurement in the classroom testing situation compared with the typical achievement testing situation. T test results, however, did not indicate any substantial difference between scores from the two different testing situations.

ACKNOWLEDGEMENTS

Love and appreciation to my parents who instilled in me my high respect for
education.

TABLE OF CONTENTS

CHAPTER	PAGE
I INTRODUCTION	1
Statement of the Problem	2
Purpose of the Study	3
Hypothesis	3
Definitions	3
II REVIEW OF LITERATURE	5
Factors Affecting Scores	5
Motivation	7
Learned Helplessness	9
Locus of Control	10
Alternatives To Standardized Tests	12
III METHODOLOGY	15
Subjects	15
Tools and Instruments	15
Data Collection	17
Data Analysis	18
IV RESULTS AND DISCUSSION	19
V SUMMARY	24
Results	24
Recommendations	25
Future Research	25
Conclusion	26
REFERENCES	27

LIST OF TABLES

TABLE	PAGE
1 SDTM Computation Subtest Scores from Greatest Loss to Greatest Gain in Percentile Rankings	20
2 SDTM Application Subtest Scores from Greatest Loss to Greatest Gain in Percentile Rankings	22

INTRODUCTION

The exclusive use of standardized tests is not necessarily the best method of assessing achievement levels of special education students. Jenkins and Pany (1978) found that if students are taught a particular reading series, they will score higher on a particular standardized test and, if they are tested with an instrument that does not match the material covered in their reading program, their scores will be significantly lower. Zigmond and Silverman (1984) indicated concern that scores, in general, are not reliable because the standardization populations include few, if any, exceptional students. The California Education Code (Special Education Resource Book, 1989) does not require the use of standardized tests in assessing students at any of the various stages from initial placement through annual and three-year reviews.

Even though facts support the opponents of standardized tests (Goodstein, 1982; Jenkins & Pany, 1978; McCarthy, 1980; Zigmond & Silverman, 1984), school districts still utilize them routinely for evaluating students. Assessment scores assist in screening, determining eligibility, program planning, monitoring student progress, and evaluating the program (Lewis & McLoughlin, 1986). Some teachers are strongly urged to include standardized test scores from at least one test for mathematics, reading, and writing. Discussion of academic achievement at the Individualized Educational Program (IEP) meeting includes these test scores. Often these test scores are translated into grade level equivalents that indicate how much or how little the student has achieved in the past year. The student is told that he is still functioning on

the third or fourth grade level in one or more of the three major academic areas of reading, mathematics, and written language. The probability exists that the student has been given this same information for the past several years. Test scores are also used to decide what type of classes the student should be placed in for the coming year. Students may well be enrolled in classes designed to remediate their deficiencies with little or no thought to the placement they would prefer or to the development of any talent they may have. Improvement of these test scores as close to grade level as possible prior to graduation is such a high priority with the educational system that it is possible that a student could be kept in remedial classes for his entire school career. Repeatedly placing the student in remedial classes in many cases will prevent the student from experiencing elective or vocational coursework.

Statement of the Problem

Standardized tests frequently utilized in the school district where this research was conducted are: the Stanford Diagnostic Test for Mathematics and Reading for annual IEPs, and the Woodcock-Johnson Psycho-Educational Battery, Part Two: Tests of Achievement for three-year evaluations. Test scores are used in making decisions about a student's life; therefore, they must be accurate. The methods used to administer the tests may cause students not to work to their full potential. Consequently, scores obtained may not be valid.

Students frequently ask why they have to take the same tests again. Regardless of the reasons given, students are not convinced that their scores on these tests have any direct influence on their education or their lives.

Purpose of the Study

The purpose of this research was to investigate whether or not students put forth their full effort when taking standardized achievement tests.

Hypotheses

Null hypothesis 1. There will be no significant difference between computation subtest scores achieved by students tested with the standard booklet of the Stanford Diagnostic Mathematics Test (SDMT) under typical achievement testing procedures and students tested with a copy of the SDMT using routine classroom testing procedures.

Null hypothesis 2. There will be no significant difference between application subtest scores achieved by students tested with the standard booklet of the Stanford Diagnostic Mathematics Test (SDMT) under typical achievement testing procedures and students tested with a copy of the SDMT using routine classroom testing procedures.

Limitations

Limitations were imposed by the small number of students tested.

Definitions

Typical testing procedure (TTP). Students are removed from classes individually or in small groups and administered a standardized achievement test.

Standardized test. A test in which the administration, scoring, and interpretation procedures are structured or set, usually norm referenced.

Classroom testing procedure (CTP). The test is worked into the regular classroom curriculum where immediate rewards can be realized. The instrument is designed to look like a personalized, teacher-made test.

Criterion-referenced tests. An informal assessment device that assesses skill mastery and compares the student's performance to external criteria.

Individual Educational Program (IEP). Written plan, including educational goals and objectives, designed to meet the educational needs of each individual eligible for special education, based on present levels of performance.

IEP team. A group comprised of an administrator, special and regular education teachers, parents, and sometimes the student, as well as any other individuals involved in providing services for the student, for the purpose of designing the student's Individualized Educational Program.

Stanford Diagnostic Test. A norm referenced, standardized achievement test used for screening, grouping, and evaluating students, and measuring curriculum effectiveness (Lewis & McLoughlin, 1986).

Woodcock-Johnson Psychoeducational Battery: Achievement Test. A norm referenced, standardized test that covers basic skills and content area subjects (Lewis & McLoughlin, 1986).

CHAPTER II

REVIEW OF LITERATURE

Each year there are approximately 200 million achievement tests administered in the United States. The majority of these tests are standardized tests (White, 1983). Data collected from these tests are most often used in making decisions concerning screening/referral, eligibility, program planning, program evaluation and monitoring student progress (Lewis & McLoughlin, 1986; Marston, 1983; Ysseldyke & Algozzine, 1982). If there is a discrepancy between test scores and a teacher's perception about a student's achievement, the test scores will generally be considered the more valid indicator of achievement (Jenkins & Pany, 1978). Information gathered by members of the IEP team through observation, regarding referral and intervention decisions, is ignored almost completely, while test scores from standardized tests are utilized in making eligibility decisions (Ysseldyke & Thurlow, 1984). According to White (1983) and Jenkins and Pany (1978), test scores are often inaccurate due to a variety of factors and could cause educators to make inappropriate decisions.

Factors Affecting Scores

Norming of standardized tests is one of the areas criticized by educators. Most test developers and publishers provide scant data on the appropriateness of their tests for use with handicapped children. Many of the 27 most widely used normative tests included no evidence to indicate their validity for use with handicapped children (Fuchs, 1985).

Method and setting for the administration of the tests affect the student's ability to do well (Burtelow, 1987). Unclear directions, lack of student preparation for the test, or a nonsupportive or anxiety-provoking atmosphere may cause scores to be lower than the students are really capable of achieving (White, 1983). Checking work, using elimination strategies, timing, and following directions are directly related to how well one will do on a test. The large number of test formats utilized, which are frequently complicated, may cause students to become confused and, consequently, to score lower. Recommendations have been made to counselors that, due to the many sources of error in group-administered tests, they should consider alternate sources of information for the slow learner (Karmos & Karmos, 1984).

Jenkins and Pany (1978) discussed the curriculum bias related to standardized reading achievement tests. The research utilized four reading achievement tests and five of the most commonly used commercial reading series. The researchers found that, if a teacher using one of the five reading series selected an inappropriate standardized test, results showed little or no student achievement. However, if a test compatible with the reading program was selected, expected or even greater than expected student growth might be achieved. Therefore, standardized tests may report inadequate achievement levels, even though the instructional program is adequate.

When one expects certain behavior of others, one is likely to act in ways that make the expected behavior more likely to occur (Rosenthal & Babad, 1985). Extensive research in the areas of student academic performance, the workplace, therapy clinics, and athletics supports this phenomenon. Research has repeatedly shown that students'

academic gains can be greatly influenced by their teachers' expectations of their potential. A student's physical appearance, as well as past academic history, can substantially influence a teacher's expectations of the student. These expectations will then, in turn, influence the success or failure experienced by the student (Rosenthal & Babad, 1985).

Motivation

Madeline Hunter (1987) defined motivation as a student's intent to learn or accomplish something. Motivation is learned; it is not genetic. Lack of motivation is a major factor in preventing school success for special education students (Licht, 1983; Mercer, 1987).

"Learning disabled adolescents often do not perform well on tests because it is boring to repeat the same task time and time again or because it is highly frustrating to be repeatedly placed in a failure situation" (Alley & Deshler, 1979, p. 31).

"Motivational processes influence a child's acquisition, transfer, and use of knowledge and skills, yet educationally relevant conceptions of motivation have been elusive" (Dweck, 1986, p. 1040).

Underachievement originally was associated with gifted students who did not perform up to their ability. Today, however, underachievement is a concern directed towards achievement in general, even among students with learning disabilities. Serious underachievers never catch up to their potential, regardless of whether they are of low, medium or high mental ability (Avazian, 1987).

Academic achievement is not solely determined by intelligence. A consistent and positive relationship has been found to exist between self-concept and academic achievement. Learning disabled students have a lower self-concept than do non-learning disabled students. Learning disabled readers scored significantly lower on the Piers-Harris Children's Self-Concept Test than did normal readers, and the self-concept scores decreased as age and grade level increased. Significant relationships have also been found between low self-esteem and mathematics, reading recognition, and general information scores for learning disabled students (Avazian, 1987).

Low-ability students have a more negative attitude toward tests than do high-ability children (LaVoie & Hodapp, 1987). Low achievers tend to be more negative about how schools utilize standardized test scores than are high achievers. These low achievers may well react negatively to standardized tests if they believe that the results of the tests will be a major factor in decisions concerning their academic programs. Studies show that student attitudes could account for up to 25% of the variance of achievement in courses or learning tasks (Karmos & Karmos, 1984). When students were asked about their attitudes towards standardized tests, "47% of them agreed that 'taking achievement tests is a waste of time' and 36% agreed that 'achievement tests are dumb'" (Karmos & Karmos, 1984, p. 61).

Learning goals are associated with individuals who seek to increase their competence and to understand or master something new, while performance goals are associated with individuals who seek to gain favorable and avoid negative judgments of

their competence (Dweck, 1986). Individuals who believe that intelligence is a fixed trait tend to lean towards performance goals, whereas individuals who believe that intelligence is a malleable quality lean towards learning goals. Research indicates that individuals who focus on performance goals will have a tendency to avoid or withdraw from a challenge (Dweck, 1986). This tendency of the goal-oriented student could have a significant effect on the accurate assessment of the student's achievement levels.

Harackiewicz and Manderlink (1984) found that performance-contingent rewards had a more positive effect on the level of motivation for low achievers than for high achievers. The use of extrinsic control techniques, such as token economies and contingency contracting, has shown some success in improving motivation towards better performance of academic skills of learning disabled adolescents (Mercer, 1987). Tokens may consist of play money or trading stamps which are collected and then exchanged for a desired object or activity. A contingency contract is a contract between a teacher and student stating that, when a desired behavior is performed, it is followed by the student being able to engage in a designated reward activity.

Learned Helplessness

Bardwell (1984) described learned helplessness as "those instances in which an organism has learned that outcomes are uncontrollable by his responses and is seriously debilitated by this knowledge" (p. 192). This theory ascribes low effort to a belief in no control (Snyder, Smoller, Strenta & Frankel, 1981), causing the individual to give up and stop trying (Miller, 1986). This belief is one of the factors affecting the level of motivation that students have towards tests or other learning tasks. Following a

two-year study of 82 children, Fincham, Hokoda and Sanders (1989) reported finding a relationship between symptoms of learned helplessness in third grade students and achievement test scores of these same students in the fifth grade. To what extent can learned helplessness affect one's ability to function at full capacity is the question. Perhaps one of the most dramatic examples was reported by Snyder et al. (1981). They stated that some of the performance deficits associated with aging may be due to learned helplessness. When institutionalized aged were given a sense of control, it served to slow down, arrest, or even reverse these deficits. Other researchers believe that "what is often passed off as retardation or an IQ deficit may be the result of learned helplessness" (Walters, 1983, p. 5).

Locus of Control

An individual who believes that his behavior affects or controls the reinforcement he receives is said to have internal locus of control. People who believe their behaviors make no difference, that good or bad things happen because of chance or the whims of powerful others, are said to have external locus of control. "A positive relationship between grades and internal locus of control has been established" (Walters, 1983, p. 2). Children with external locus of control who attribute past failures to uncontrollable causes, such as ability or luck, will not persist in the presence of failure. According to Tomlinson (1987), learning disabled students are found to have external locus of control more often than internal locus of control (Dudley-Marling, Snider, & Tarver, 1982). If an individual perceives that he lacks control, this perception will

cause negative feelings that will affect the individual's performance, motivation, and self-esteem (Walters, 1983).

Dickens and Perry (1982) studied the effects of perceived control on college student achievement and attributions. Results indicated that 10% of the students in the program perceived themselves to have little or no control over their academic lives. These students spent less time studying, attended fewer classes, and felt it was not important to do well academically. Field studies following the laboratory procedures revealed that perceptions of control were related to attributions about course outcomes and even affected classroom behavior. Students who had a higher perception of control made more internal attributions and took a greater responsibility for their academic outcomes (Dickens & Perry, 1982).

Individuals who determine that the reason for an event's outcome is lack of ability may feel that the situation is hopeless and expect to fail again. These individuals will study less and put forth less effort on subsequent tasks. Such are the assumptions of the attribution theory in relationship to motivation (Wilkinson & Sanchez, 1985).

"Private self-consciousness (PSC) refers to the dispositional tendency to be attentive to the private, covert aspects of oneself" (Franzoi, 1986, p. 1). Individuals with low PSC will be less willing to discover something negative about themselves than will individuals with high PSC. In other words, students with low PSC have a greater need for self-defense of their self-esteem than for self-knowledge. One might conclude from this assertion that a student with low PSC will do whatever is necessary to protect self-esteem.

Egotism theory suggests that a person will persist more on tasks considered to be difficult and will demonstrate less persistence on tasks that are considered easy or moderately difficult (Frankel & Snyder, 1987). Research studies on depressed college students revealed that, as tasks were described as increasingly more difficult, the students put forth more effort on the tasks. The researchers concluded that the subjects felt that they did not have to worry about the implications of failure when the task was described as extremely difficult as much as they had to worry about failure when the task was described as easy or moderately difficult.

One major theme that permeates literature on motivating special education students is that of involving students in all aspects of the curriculum. Specifically, students should be involved in selecting and ordering their own academic goals (Mercer, 1987).

Alternatives To Standardized Tests

Direct repeated measurement. Utilization of direct and repeated measurement of academic skills is the most satisfactory method of measuring student progress toward IEP goals (Marston, 1982). This method has also proved to be as effective as commercial measures of achievement in predicting student eligibility. Direct measurement of academic skills has been used successfully to distinguish between learning disabled and low-achieving students, whereas standardized tests were unable to make this differentiation.

Continuous-monitoring model. Non-standardized or locally-normed, direct measures of student performance in specific academic or social adjustment areas are the measurements involved in continuous-monitoring methodology. These measures will

generally be supplemented by informal observation, review of classwork, and parental input (Ysseldyke & Algozzine, 1982).

Informal assessment. Informal testing focuses on student achievement in relation to the demands of the environment and the task rather than to a norm group (Zigmond & Silverman, 1984). This procedure also allows a teacher to evaluate samples of student behavior in relation to specific instructional concerns (Zigmond & Silverman, 1984). Informal tests measure what is taught and do not promote the teach-to-the-test situation. Informal assessments must be planned carefully, administered systematically, and interpreted precisely. They cannot be done in a casual, haphazard manner. The more often assessment is done, the more likely a teacher will make appropriate program and instructional changes leading to a more individualized educational program. Informal assessment can be done daily, whereas standardized testing cannot.

Criterion-referenced test. "A criterion-referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specific performance standards" (Goodstein, 1982, p. 38). Jenkins and Pany (1978) indicated that educators need an instrument to measure learning that is sensitive to curriculum differences, such as a criterion-referenced or curriculum-based assessment system. The teacher-made criterion-referenced tests can ensure a maximum fit between what is taught and what is tested. Criterion-referenced tests not only measure precisely the skills which have been taught, but they often utilize the same response mode that was used during the instructional phase (Zigmond & Silverman, 1984).

Functional level testing. Students are tested at the level on which they are functioning and being instructed instead of being tested at their particular grade level (Archer, 1984). Four issues that support functional level testing are: motivation, program evaluation, reliability of test scores, and curriculum match (Archer, 1984).

When a student is given a test that is too hard, it can be expected that he will be motivated to quit responding or mark his answers randomly. If the test is too easy, he may become careless with his answers. Standardized tests are designed to measure skills or abilities at a specific level. A test that is too easy or too hard offers too few items from which to judge a student's academic ability. "Scores for special education and disadvantaged students are often unstable" (Archer, 1984, p. 4).

Gains made by students being instructed at their developmental level often are not measured by in-level testing if the test level is inappropriate (Archer, 1984). Very little diagnostic information is obtained when a student is tested at one level while being taught at another.

Carver (1974) emphasized the need for technically adequate procedures to evaluate within-individual gain to replace norm-referenced methods if pupil progress is to be validly indexed. Such a measurement system, as outlined by Jenkins, Deno, and Mirkin (1979) would adhere to the following guidelines: the system must be relevant to the students' curriculum, sensitive to growth, flexible and adaptive to various instructional objectives, repeatedly administrable, and easily administrable. Direct measurement techniques come closest to fulfilling these characteristics (Marston, 1983).

CHAPTER III

METHODOLOGY

High school special education students may be inappropriately placed in classes due to scores they receive on standardized achievement tests (Jenkins & Pany, 1978; White, 1983). Students may be unable to see any value in the tests and may not work up to their full potential when taking the tests. Scores from these tests are routinely used by IEP teams to determine a student's achievement level and ultimately to assist the team in academic placement of the student. The purpose of this research was to investigate whether special education students tested in the standard manner score differently from the same students tested in a classroom situation.

Subjects

Twenty-one learning disabled, 9th through 11th grade students enrolled in mainstream Mastery Math classes, identified and placed in special education while in elementary school, comprised the subjects of this study. There were 13 males and 8 females.

Tools and Instruments

The Stanford Diagnostic Mathematics Test, Blue Level, Form G, was selected for this research for two reasons: (a) it was routinely utilized by special education teachers in this particular school district, at the time of this study, for assessing student achievement; and (b) it was originally developed with the low-achieving student in mind. Most achievement tests are designed to measure the broad range of ability of

pupils in specified grades by including about the same number of easy and difficult questions. The SDMT was designed to allow even low-achieving students the opportunity to experience some success by asking more easy questions than difficult ones. It is possible to use the SDMT as an achievement test, even though it was primarily designed as a diagnostic tool, because it is norm-referenced. The raw scores can be translated into percentile ranks, stanines, and grade equivalents (Beatty, Madden, Gardner, & Karlsen, 1976).

Two subtests, Computation and Application, from the Stanford Diagnostic Mathematics Test (SDMT) (Beatty et al., 1984) were used for testing students. Problems in multiplication, division, addition, and subtraction of whole numbers, fractions, and decimal numbers make up the SDMT Computation subtest. Word problems utilizing computation skills, skills in graph and chart reading, and basic geometry, are covered by the SDMT Application subtest. Problems on both subtests are multiple choice. Administration procedures for the SDMT do not allow using calculators or writing in the test booklet. Students may use scratch paper to compute the problems and then must select their answers on a machine-scored answer sheet.

Teacher-made tests based on the problems from the SDMT Computation and Application subtests were used for the first testing session. Problems from the two subtests were xeroxed out of the standard test booklet in order to compile teacher-made test booklets. The teacher-made booklets contained only the problems on plain white paper. No instructions or examples were included. Type size and style, as well as arrangement on the pages, remained the same from the standard booklet to

teacher-made booklets. All standardized requirements for administering the SDMT according to the test manual were followed with the teacher-made tests.

Data Collection

Approximately six weeks after school started, the first testing session took place. This time was selected because it was the end of the first grading period. The Classroom Testing Procedure (CTP) took place in the regular classroom with both regular and special education students taking the tests. The students were informed that the test was their regular end-of-grading-period examination and that their scores would be used in computing their grade for the current grading period. In order to make the test appear more routine and to motivate the students to work to their full potential, a class competition was introduced. The class scoring the most points on the two subtests was to receive a pizza party during their regularly scheduled lunch hour.

On two consecutive days, teacher-made tests based on the two subtests of the SDMT were administered in accordance with the Stanford Diagnostic Test requirements. Machine-scored answer sheets were employed and time constraints were carefully followed. Students were permitted to use scratch paper to compute problems, but no other aids (such as calculators) were permitted. All tests were scored automatically.

Testing was done again starting in late January. The Typical Testing Procedure (TTP) for the second round of testing was the one routinely used at this particular high school when preparations were being made for the annual IEP meetings. Students were removed from classes either individually or in groups of up to six and tested using the

official SDMT Computation and Application subtest booklets. Administration was carried out in strict accordance with SDMT test requirements over a period of two days. Students were informed that the test scores would be used during their IEP meeting to help determine class schedules for the next year.

Data Analysis

Raw scores from the tests administered in October and January were collected and converted into percentiles. Percentile scores are dependent upon raw score, student's grade level, and the time of year the test is administered. October scores were based on the SDMT fall norms while January scores were based on the spring norms. Only scores from students completing both subtests during both testing procedures were used. A comparison was made of these scores for each individual student tested. Test scores were also evaluated using the standard error of measurement (SE) for the SDMT and with the t-test for repeated measures.

CHAPTER IV

RESULTS AND DISCUSSION

When given the same test problems, students would be expected to score higher on a test given after 12 weeks of instruction in basic math skills than they had on a test given prior to instruction. Students are expected to score higher on posttests than on pretests. Consequently, one would expect students to score higher on a test taken in January than on the same test taken the previous October. This research compared scores of 21 learning disabled students using the Typical Test Procedure (TTP) in January with those same students tested the previous October using the Classroom Test Procedure (CTP).

Results of the Computation subtest of the Stanford Diagnostic Mathematics Test administered in October and January are presented in Table 1. Scores were ranked from greatest percentile loss (1) to greatest percentile gain (21). Of the 21 students tested on the Computation subtest of the SDMT, 57% had a lower percentile ranking in January with the Typical Test Procedure (TTP) than they had on the same test administered the previous October with the Classroom Test Procedure (CTP).

The largest decrease in raw score from October to January was 12 points on the Computation subtest. One student's score dropped from 42nd percentile on the October test to the 27th percentile on the January test. The greatest gain on the Computation subtest was 6 points on the raw score, a gain from the 39th percentile in October to the 44th percentile in January. Twelve students decreased their percentile

Table 1

SDMT Computation Subtest Scores from
Greatest Loss to Greatest Gain in Percentile Rankings

N	Grade	October		January		%ile Difference
		Raw Score	%ile	Raw Score	%ile	
1	10	28	18	16	3	-15
2	9	34	42	32	27	-15
3	9	32	36	30	22	-14
4	11	31	19	23	6	-13
5	9	30	31	28	18	-13
6	9	20	12	10	1	-11
7	9	25	20	22	9	-11
8	9	26	22	26	15	-7
9	9	16	7	9	1	-6
10	10	21	8	15	3	-5
11	10	23	10	20	7	-3
12	9	11	2	11	1	-1
13	10	22	9	22	9	0
14	10	22	9	22	9	0
15	10	16	3	16	3	0
16	9	14	4	16	4	0
17	10	21	8	22	9	+1
18	10	22	9	23	10	+1
19	10	29	21	31	42	+3
20	9	21	13	27	16	+3
21	9	33	39	37	44	+5

ranking from October to January on the Computation subtest, 4 students remained the same, and 5 students increased their rank.

On the application subtest represented in Table 2, 48% achieved a lower percentile ranking in January than they had on the October test.

Application subtest results indicated the greatest decline in raw score was 8 points, from the 34th percentile in October to the 11th percentile in January. The greatest improvement shown by any student from October to January was on the Application subtest. The student gained 7 points on the raw score and advanced from the 3rd to the 18th percentile.

Test reliability, the consistency of results from one testing session to another, is often expressed in terms of standard error of measurement in raw scores. The standard error of measurement (SE) is the amount that individual test scores may vary due to chance. The SE for the SDMT Computation Subtest is + or - 2.7 points for grades 9 and 10 and + or - 2.4 points for grade 11 (Beatty et al., 1976).

Of the 21 students tested in October on the SDMT Computation Subtest with CTP, after 12 weeks of further instruction and being tested with TTP instead of CTP, 57% scored within 1 SE. Twenty-nine percent achieved test scores in the second testing lower than 1 SE. Only ten percent of the students showed improvement on the second testing greater than 1 SE.

The SE for the SDMT Application subtest is + or - 2.4 for grades 9 and 10 and + or - 2.3 for grade 11 (Beatty et al., 1976). Once again, of the 21 students tested in October on the SDMT Application subtest with CTP, after 12 weeks of further

Table 2

SDMT Application Subtest Scores from
Greatest Loss to Greatest Gain in Percentile Rankings

N	Grade	October		January		%ile Difference
		Raw Score	%ile	Raw Score	%ile	
1	10	18	34	13	11	-23
2	10	15	21	7	1	-20
3	10	12	10	7	1	-9
4	10	13	13	10	5	-8
5	10	13	13	10	5	-8
6	9	17	34	16	26	-8
7	11	12	7	10	3	-4
8	10	10	8	9	4	-4
9	9	10	6	9	3	-3
10	9	13	17	13	14	-3
11	9	8	4	9	4	0
12	10	9	4	10	5	+1
13	10	22	56	24	58	+2
14	10	9	6	11	8	+2
15	9	10	8	12	11	+3
16	9	12	14	14	18	+4
17	9	9	4	11	8	+4
18	9	8	4	11	8	+4
19	9	6	2	10	6	+4
20	9	5	1	11	8	+7
21	9	7	3	14	18	+15

instruction and being tested with TTP instead of CTP, 52% scored within 1 SE.

Twenty-four percent scored lower on the second test than 1 SE and twenty-four percent showed improvement on the second test greater than 1 SE.

The t test for repeated measures indicated no significant changes in test scores from the October testing to the January testing on either of the subtests. The Application subtest required a t of + or - 2.086 or larger for a .05 level of significance. The test scores from the students tested in this program indicated a $t(20)$ of .03, $p > .05$. The Computation subtest required a t of + or - 2.086 or larger for a .05 level of significance. Test scores from the students tested for this research indicated $t(20)$ of -1.75, $p > .05$.

CHAPTER V

SUMMARY

The purpose of this study was to examine the possibility that a significant number of students do not perform up to their potential and, in fact, may perform substantially below their potential when taking routine achievement tests. No attempt was made to decide why students do not perform up to their potential on standardized achievement tests.

Learning disabled students in grades 9 through 11 who participated in this study had been placed in special education classes while in elementary school. At the time of this research, they were all enrolled in mainstream Mastery Math.

The Computation and Application subtests from the Stanford Diagnostic Mathematics Test and teacher-made copies of the same two subtests were administered. Students were tested with the teacher-made version of the two subtests in a regular classroom testing environment (CTP) and again 12 weeks later in a typical achievement test setting (TTP) with the publisher's standard test booklets.

Results

Fifty-seven percent of the students scored lower on the Computation subtest and 48% scored lower on the Application subtest when using the TTP than they did on the same subtests using the CTP. Therefore, a substantial number of students did score lower when they were tested with a standardized achievement test in a typical testing environment than they scored in a classroom testing situation.

The t test for repeated measures indicated there was no significant difference between the scores on subtests administered in October with the classroom testing procedure and scores on subtests administered in January with the typical testing procedure. The standard error of measurement applied to the raw scores for the Stanford Diagnostic Mathematics Test showed that approximately 25% of the students scored lower than 1 SE in January on both the Computation and Application subtests.

Recommendations

Educators should evaluate their use of standardized achievement test scores for writing Individual Education Programs. IEP teams need to be aware that achievement test scores may not be an accurate measurement of the ability of some of their students. Educators need to use a variety of evaluation tools for special education students. Educators should rely heavily on the observation of practical application of material taught and criterion-referenced tests. These methods, used in conjunction with the standardized or norm-referenced tests, will provide more accurate information about a student's ability.

Future Research

Researchers should examine use of achievement tests with regular education students. Some of the same situations and results may exist when regular education students take achievement tests. Perhaps standardized achievement tests are not giving an accurate evaluation of academic accomplishment in schools.

Conclusion

Results from this research were not conclusive. The t test indicated that there was no significant difference in the test scores of students tested in the CTP and the TTP. However, the fact that 25% of the students exposed to the TTP did score lower than 1 standard error of measurement, after 12 weeks of additional instruction, should not be overlooked. The t test deals with the average of a group. Special Education deals with the individual. It is the responsibility of the IEP team to provide each individual special education student with the best possible educational opportunities. The IEP team must be aware of the importance of utilizing a variety of evaluation procedures.

REFERENCES

- Alley, G., & Deshler, D. (1979). Learning strategies: Teaching the learning disabled adolescent: Strategies and methods. Denver, CO: Love Publishing Company.
- Archer, E. L. (1984). Achievement testing of the handicapped student: One school district's experience. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 243 945).
- Avazian, K. L. W. (1987). The effects of learning disabilities on a child's self-concept. Doctoral research paper, Biola University. (ERIC Document Reproduction Service No. ED 290 281).
- Bardwell, R. (1984). Failure: facilitating or debilitating. Journal of Experimental Education, 52(4), 192-194.
- Beatty, L. S., Madden, R., Gardner, E. F., Karlsen, B. (1976). Stanford Diagnostic Mathematics Test Manual for Administering and Interpreting. (Blue Level). New York: Harcourt Brace Jovanovich, Inc.
- Burtelow, J. D. (1987). Accountability rests on a not too sharp #2 pencil. English Journal, 76(3), 38-39.
- Carver, R. P. (1974). Two dimensions of tests: Psychometric and edumetric. American Psychologist, 29, 512-518.

- Dickens, W. J., & Perry, R. P. (1982). Perceived control in the college classrooms: The impact of student and teacher characteristics. Paper presented at the International Congress of Applied Psychology, Edinburgh, Scotland. (ERIC Document Reproduction Service No. ED 227 736).
- Dudley-Marling, C. C., Snider, V., & Tarver, S. G. (1982). Locus of control and learning disabilities: A review and discussion. Perceptual and Motor Skills, 54, 503-514.
- Dweck, C. (1986). Motivational processes affecting learning. American Psychologist, 41(10), 1040-1048.
- Fincham, F. D., Hokoda, A., & Sanders Jr., R. (1989). Learned helplessness, test anxiety, and academic achievement: A longitudinal analysis. Child Development, 60, 138-145.
- Frankel, A., & Snyder, M. L. (1987). Egotism among the depressed: When self-protection becomes self-handicapping. Paper presented at the Annual Convention of the American Psychological Association. New York. (ERIC Document Reproduction Service No. ED 289 120).
- Franzoi, S. L. (1986). A motivational explanation of private self-consciousness. Paper presented at the Annual Convention of the American Psychological Association. Washington, DC. (ERIC Document Reproduction Service No. ED 281 085).
- Fuchs, D. (1985). Norm-referenced tests: Are they valid for use with handicapped children? Tennessee. (ERIC Document Reproduction Service No. ED 265190).

- Goodstein, H. A. (1982). The reliability of criterion-referenced tests and special education: Assumed versus demonstrated. Journal of Special Education, 16(1), 37-47.
- Harackiewicz, J. M., & Manderlink, G. (1984). A process analysis of the effects of performance-contingent rewards on intrinsic motivation. Journal of Experimental Social Psychology, 20, 531-551.
- Hunter, M. (1987). Increasing their motivation Part I and II. Mastery teaching (pp. 11-15, 19-23). El Segundo, CA: TIP Publications.
- Jenkins, J. R., Deno, S. L., & Mirkin, P. K. (1979). Measuring pupil progress toward the least restrictive alternative. Learning Disability Quarterly, 2, 81-92.
- Jenkins, J., & Pany, D. (1978). Standardized achievement tests: How useful for special education? Exceptional Children, 49, 448-453.
- Karmos, A. H., & Karmos, J. S. (1984). Attitudes toward standardized achievement tests and their relation to achievement test performance. Measurement and Evaluation in Counseling and Development, 17(2), 56-66.
- LaVoie, J. C., & Hodapp, A. F. (1987). Children's subjective ratings of their performance on a standardized achievement test. Journal of School Psychology, 25, 73-80.
- Lewis, R., & McLoughlin, J. (1986). Assessing special students (2nd ed.). Columbus, Ohio: Merrill Publishing Company.
- Licht, B. G. (1983). Cognitive-motivational factors that contribute to the achievement of learning disabled children. Journal of Learning Disabilities, 16(8), 483 - 490.

- Marston, D. (1982). Predictive efficiency of direct, repeated measurement: An analysis of cost and accuracy in classification. (Report No. IRLD-RR-104). Washington, DC: Department of Education. (ERIC Document Reproduction Service No. ED 226 052).
- Marston, D. (1983). A comparison of standardized achievement tests and direct measurement techniques in measuring pupil progress. (Report No. IRLD-RR-126). Washington, DC: Office of Special Education and Rehabilitative Services. (ERIC Document Reproduction Service No. 236 198).
- McCarthy, M. M. (1980). Minimum competency testing and handicapped students. Exceptional Children, 47(3), 166-172.
- Mercer, C. (1987). Adolescents and adults. In V. Knight (Ed). Students with learning disabilities (pp. 279-80). Columbus, OH: Merrill Publishing Company.
- Miller, A. (1986). Performance impairment after failure: Mechanism and sex differences. Journal of Educational Psychology, 78(6), 486-491.
- Rosenthal, R., & Babad, E. Y. (1985). Pygmalion in the gymnasium. Educational Leadership, 43(1), 36-39.
- Snyder, M. L., Smoller, B., Strenta, A., & Frankel, A. (1981). A comparison of egotism, negativity, and learned helplessness as explanations for poor performance after unsolvable Problems. Journal of Personality and Social Psychology, 40(1), 24-30.
- Special Education Resource Book (11th ed.). (1989, December). (Available from Sprague Associates, P.O. Box 918, Woodbridge, CA 95258).

- Tomlinson, L. M. (1987). Locus of control and its affect on achievement. (ERIC Document Reproduction Service No. ED 276965).
- Walters, E. (1983). Verbal self-instruction for the mentally retarded: The missing link. Paper presented at the Annual International Convention of The Council for Exceptional Children. Detroit, MI. (ERIC Document Reproduction Service No. ED 230 016).
- White, K. (1983). An evaluation of training in standardized achievement test taking and administration. Final report of the 1981-82 Utah state refinements to the ESEA title I evaluation and reporting system. (Contract No. 300810 271). Salt Lake City, Utah: Utah State Office Education. (ERIC Document Reproduction Service No. ED 238 931).
- Wilkinson, C. Y., & Sanchez, J. E. H. (1985). Attribution theory, learned helplessness and achievement: Implications for the exceptional Hispanic child. (Contract No. 300-83-0272). Washington, DC, Office of Special Education and Rehabilitative Services. (ERIC Document Reproduction Service No. ED 261 486).
- Ysseldyke, J., & Algozzine, R. (1982). Critical issues in special and remedial education. Boston: Houghton-Mifflin, 123-151.
- Ysseldyke, J. E., & Thurlow, M. L. (1984). Assessment practices in special education: Adequacy and appropriateness. Educational Psychologist, 2(3), 123-136.

Zigmond, N., & Silverman, R. (1984). Informal assessment for program planning and evaluation in special education. Educational Psychologist, 19(3), 163-171.