University of Nebraska at Omaha
DigitalCommons@UNO

Student Work

7-1-1996

A Group Achievement Test as a Predictor of Support Services Beyond the Classroom

Deborah D. White
University of Nebraska at Omaha

Follow this and additional works at: https://digitalcommons.unomaha.edu/studentwork

Recommended Citation
White, Deborah D., "A Group Achievement Test as a Predictor of Support Services Beyond the Classroom" (1996). *Student Work*. 2646.
https://digitalcommons.unomaha.edu/studentwork/2646

This Thesis is brought to you for free and open access by DigitalCommons@UNO. It has been accepted for inclusion in Student Work by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.

A GROUP ACHIEVEMENT TEST AS A PREDICTOR OF SUPPORT SERVICES BEYOND THE CLASSROOM

An Ed. S. Field Project

Presented to the

Department of Psychology

and the

Faculty of the Graduate College

University of Nebraska

In Partial Fulfillment

of the Requirements for the Degree

Specialist in Education

University of Nebraska at Omaha

by

Deborah D. White

July 1996

UMI Number: EP74190

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

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

UMI EP74190
Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ED.S. FIELD PROJECT ACCEPTANCE

Acceptance for the faculty of the Graduate College, University of Nebraska, in partial fulfillment of the requirements for the degree Specialist in Education, University of Nebraska at Omaha.

Committee

Name	Department/School
Lisa Kelly-Vance	Psychology
Robert Henley Woody	Psychology
Joseph C. LaVoie	Psychology
Thomas C. Lovitt	Special Educ. & Com. Dis.

Chairperson Lisa Kelly-Vance

Date July 15, 1996

Acknowledgements

I would like to thank Dr. Lisa Kelly-Vance for serving as committee chairperson, advocate, enthusiast, and supporter during this research project. Appreciation is extended to Dr. Joseph LaVoie, Dr. Thomas Lorsbach, and Dr. Robert Woody for serving as committee members and mentors throughout the School Psychology program.

My gratitude is also extended to the school district for granting permission to conduct this research project, and the school psychologists who collected the data, shared in my enthusiasm, and offered their encouragement. Special recognition is given to Lynne Newville for her brainstorming assistance during the initial stages of this project, and to Evelyn Headen for her support, guidance, and counsel throughout this entire endeavor.

Foremost, special thanks go to my children, Nicole and Travis, for their years of unyielding acquiescence and personal sacrifices while I have been a graduate student, and especially to my parents for their unconditional support not only during graduate school, but throughout my lifetime.

Table of Contents

List of Tables

Abstract

The predictive value and relative accuracy of a student's performance on a group achievement test, Metropolitan Achievement Test-Seventh Edition (MAT7), was contrasted with participation or nonparticipation in support services options beyond informal interventions implemented by classroom teachers. Results indicated that a student's performance on the MAT7 significantly discriminated between the Participation and Nonparticipation Groups. The Total Reading, Total Math, and Basic Battery scores significantly contributed to the total variance for group membership (Participation and Nonparticipation Groups). Separate analyses of the National and Local Percentile scores found that the National scores accounted for 25 percent of the variance and the Local scores accounted for 20 percent of the variance. Overall classification accuracies yielded approximately 78% for National percentile scores and approximately 71% for the Local percentile scores. Chi-square analyses found that the three broad scores at or below one standard deviation of the mean significantly increased the likelihood that students received additional support services. Cutoff scores distinguishing the types of supportive services were established for the individual broad scores at either one standard deviation or 1 1/2

standard deviations below the mean for most of the Local and National percentile scores. Based on the results from this population sample, the use of group achievement test data for the initial and efficient identification of students who may be academically at-risk is recommended. Suggestions for future research include the investigation of additional factors not only contributing to the amount of variance distinguishing students who are academically at-risk, but also factors affecting the misclassification of students.

Chapter 1

Introduction and Statement of Purpose

Students whose academic performances are significantly lower than their peers are much more readily identified as students in need of additional instruction and remediation. However, students who may not be as easily distinguishable may also be in need of further assistance. Frequently, these students are considered at-risk for academic failure, but are not identified due to the undifferentiated area between the norm and the lower extreme end of the learning continuum. The purpose of the present study is to investigate the degree to which a student's performance on a group achievement test is indicative of an academic concern warranting further investigation.

Researchers continue to investigate strategies that can reliably, efficiently, and promptly identify the population of children who are academically at-risk. Students who are not achieving at a rate commensurate with their ability may be identified by either a person familiar with the student's current level of academic performance or by cursory examinations (i. e., screening programs) (Salvia & Ysseldyke, 1995).

Referral decisions to seek further assistance can be initiated by parents, the students, or other individuals, such as teachers, school psychologists, and administrators

(Salvia & Ysseldyke, 1995). However, most students are dependent on teacher referral for further academic support or assessment. Only 73 percent of students who are referred by teachers are placed in special education classes. Salvia and Ysseldyke (1995) postulate that this somewhat low "hit rate" (proportion of accurate positive decisions) is a result of tolerance variability of acceptable behavior by different teachers.

Screening and child find procedures frequently target preschool or kindergarten entry. Ogden and Germinario (1988), however, discuss the need for a perspective that extends from kindergarten through the twelfth grade. A student's progression along the learning continuum may be affected by variables that impede or reduce learning at any time. Since identification of students who are having academic difficulties is beneficial for the student, school system, and society, screening methods need to continue beyond kindergarten. The periodic administration of group achievement tests as a general screening procedure after kindergarten may effectively identify those students who warrant further investigation in later grades.

The assessment literature and test publishers routinely discuss group achievement tests as a viable screen by which students' academic strengths and weaknesses may be diagnostically interpreted to provide appropriate

instruction and resources (Salvia & Ysseldyke, 1995). Further collaborative evidence provided by the teacher and parents may determine the degree of academic risk that would warrant further assessment or the prompt implementation of support services. In this respect, the student's performance on the achievement test becomes the initial signal for further investigation of additional remediation. Students are no longer solely dependent on the teacher for identification.

Group achievement testing has become the norm across the country (Hansen, 1993; Mitchell, 1992). However, the rampant administration of group achievement batteries has not been for screening, but for establishing a school system's accountability (Mitchell, 1992). The number of group achievement test administrations within the American schools is estimated to be from over one million per day (Lyman, 1991) to 250 million per year (Salvia & Ysseldyke, 1995).

Testing materials are costly for school systems, surpassing $24 million for achievement testing alone (Mitchell, 1992). Since it is logical to assume that group achievement testing will continue, using the existing test results to screen students who are academically at-risk would be a cost-efficient procedure for school systems and society.

School districts that administer group achievement tests on a regular basis have, at their disposal, valuable information pertaining to the academic achievement of their students. Given the controversial philosophies driving current practices and the prominence of achievement testing in American schools, sound studies are essential for the validation (or nonvalidation) of group achievement batteries as a method of initially identifying students who are academically at-risk. Examining achievement testing data may help to alert school professionals to students who may be having academic difficulties and are potentially at-risk for academic failure (Salvia & Ysseldyke, 1995). Yet to be empirically investigated is the degree to which a student's performance on a group achievement test is indicative of a concern warranting further collaboration. The purpose of this study is to determine whether group achievement test scores can accurately identify students who are academically at-risk and in need of additional support services beyond the classroom.

Review of the Relevant Literature

Students Who Are At-Risk For Academic Failure

The research literature has yet to define precisely students who are at-risk. In 1987, the Council of Chief State School Officers listed over 67 characteristics associated with the at-risk student, including disciplinary

problems, lower achievement, inadequate parental support, and physical problems. A small, but growing body of research has identified characteristics associated with at-risk, low-achieving students, such as academic difficulties, low socioeconomic status, lack of structure, inattentiveness, distractibility, low self-esteem, social and family stress, unsupportive home, poor attitude toward school, discipline problems, fear of failure, and lack of motivation (Lehr & Harris, 1988; Mitman, 1985; Rumberger, 1987). Bay and Bryan (1992) listed low achievement and poor work-related behaviors (i.e., inattentiveness, disorganization, noncompliance) as the salient characteristics that are associated with students who are academically at-risk. Further, the investigators purported these characteristics as the primary reasons for referral for special education services. Based on the results of third grade data, reading achievement, intellectual ability, retention, and socioeconomic status were found to differentiate significantly high school dropouts from graduates (Lloyd, 1978).

A meta-analysis of the literature by Rumberger (1987) revealed that the most prominent characteristics commonly identified were poor academic history, poor test performance, and family socioeconomic status. Using these characteristics as criterion measures, Payne and Payne

(1991) investigated the ability of teachers to identify academically at-risk students and found significant correlations between teacher classifications and all three criterion variables. Of particular interest was the use of grade retention to define poor academic history and a group achievement test, the Iowa Test of Basic Skills, as their measure of poor test performance. Overall, concurrence on the characteristics associated with at-risk status has yet to be accomplished. After Payne and Payne's (1991) investigation, they offered a simplistic, but possibly a realistic, definition of students who are at-risk as "learners who are not achieving up to their potential or are not meeting teacher expectations." (p. 116)

Agreement exists for the early identification of students who are academically at-risk. Further, the purpose of early identification is to implement responsive interventions addressing educational difficulties that could potentially decrease or entirely alleviate the initial academic weakness(es) impeding success (Lehr & Harris, 1988).

Accountability, ethics, and legal issues. Paralleling the individual's right for early identification of academic difficulties is a concern regarding the negative impact that many of these individuals have on society when their educational needs are neglected. School systems are now,

more then ever, being pressured by society to meet the ongoing and ever-changing educational needs of individual students. Close attention to American school systems has stemmed from: (1) publications, such as A Nation at Risk (U.S. National Commission on Excellence in Education, 1983) and Crossroads in American Education (National Assessment of Educational Progress, 1989); (2) federal legislation, such as Individuals With Disabilities Education Act (U. S. Department of Education, 1995), Rehabilitation Act: Section 504 (Martin, 1994), and Goals 2000: Educate America Act (U. S. Department of Education, 1990); (3) case law pertaining to children's educational rights (Nurcombe & Partlett, 1994); and (4) various interest groups (e.g., taxpayers, parents of children with specific disabilities).

Another consideration for identifying students who are at-risk for academic failure is the ethical responsibility set forth by professional organizations (e.g., National Association of School Psychologists and American Psychological Association). Also, ethical standards are often directly affected by legal influences that further dictate professional practices (i.e., advocating for the best interest of the child, service delivery) (Prasse, 1990). Although most educators recognize the importance of early identification and interventions for reversing the long-lasting, insidious effects of school failure, efficient

and reliable methods for detecting students who are academically at-risk have yet to be empirically identified. School systems need to strive for better methods of identifying students who are academically at-risk.

Current methods of identifying students who are academically at-risk. Teacher referrals, teacher rating scales, and direct questionnaires have been the most prominent methods of identifying the academically at-risk population (Gresham, Reschly, & Carey, 1987). An extensive review of the literature revealed only three studies that have used current group achievement tests as a criterion variable for establishing correlations with the teacher data. Methods of investigation generally consisted of the teachers first identifying the students whom they deemed as academically at-risk and, then various criterion measures (i.e., the characteristics associated with students who are academically at-risk) were used for comparison.

Payne and Payne (1991) found that teachers are moderately reliable predictors (r = .40) of students' at-risk status. Scores on the Iowa Test of Basic Skills (Reading score, Composite score) were moderately correlated (r = .46, r = .40, respectively) with teacher identification. The investigators' findings provide support for group achievement testing results as relevant information regarding the identification of academically at-

risk elementary students. Further, Payne and Payne (1991) support the position that classroom teachers are invaluable resources for both substantiating or discounting the information provided by group achievement tests.

Pedulla, Airasian, and Madaus (1980) investigated the accuracy of teachers' ratings of students on I. Q., English and math performance, and 12 social and academic behaviors related to school success. The results yielded correlations which suggested that teachers' judgments of students' performances "tap a dimension similar to that tapped by standardized tests" (Pedulla, et al., 1980, p. 307). A second factor pattern revealed two academic behaviors (i.e., attention span and persistence) also have a correlational relationship with the standardized test data. Intuitively, the investigators were not surprised with these results, explaining that "teachers cannot disentangle the purely cognitive behaviors from others [behaviors observed on a daily basis] that are related though not strictly of a cognitive nature" (p. 307).

Data supplied by Lloyd (1978) also suggest that, as early as third grade, students differ significantly on factors associated with at-risk status. Reading achievement measured by a group achievement test, the California Achievement Test, was considered the strongest predictor (r = .37) at both the third and sixth grades for later high

school success or failure. Lloyd (1978) concludes that "early indications of later school difficulty should be helpful in working with individual students in an attempt to change what appears from research to be an inevitable outcome" (p. 1,199).

Standardized Achievement Tests

General description. Achievement tests are intended to measure the results of previous learning. An achievement battery is a collection of subtests that sample different subject areas; typically addressed are vocabulary, reading, mathematics, spelling, language, science, social studies, and writing. Standardized achievement tests are prepared commercially by specialists in measurement and subject matter, with carefully constructed uniform and standardized procedures (e.g., the same set of questions, directions, time constraints, scoring).

Standardized, norm-referenced tests compare a person's performance to the performance of other members in a given group (Lyman, 1991; Witt, Elliott, Kramer, & Gresham, 1994). The data from norm-referenced tests yield a student's relative standing along a continuum of attainment in comparison to groups established at national and local levels.

Group administrations are considered an efficient tool for the cost-effective collection of information regarding a

student's performance (Witt et al., 1994), as entire classes can be given the test in a relatively short period of time. Objective tests also reduce scoring errors, and permit uniform, easy, and rapid scoring. The answer sheets can be scored either by properly trained clerks or by computers, further reducing the amount of time that school professionals must devote to group achievement testing.

Anastasi (1982) asserts a distinct advantage for group achievement batteries administered on a regular basis:

> They permit horizontal or vertical comparisons, or both. Thus, an individual's relative standing in different subject-matter areas or educational skills can be compared in terms of a uniform normative sample [vertical comparison]. Or the child's progress from grade to grade can be reported in terms of a single score scale [horizontal comparison]. (p. 400)

Although group tests have many desirable features, there are some disadvantages that need to be considered when interpreting the data, such as reduced opportunities to establish rapport, to solicit cooperation, and to maintain the interest of examinees. Critics of group achievement tests (as cited in Anastasi, 1982) also allude to the simplicity of multiple-choice items as hindering the measurement of higher level thinking skills that are a priority in today's curriculum. Anastasi (1982), however,

discounts this criticism based on both empirical evidence and the current test construction practices performed collaboratively by well-trained psychometricians, psychologists, and educators. She argues that multiple-choice items "can be written to tap complex thinking processes, reasoning, evaluation of arguments, and the application of knowledge to new situations" (p. 398).

In summary, a word of caution must be given when interpreting only low scores on a group achievement test, and a best practice approach would dictate further collaboration of other relevant information beyond that specific test setting. Lyman (1991) purports that the abolishment of testing is not the answer for schools, but rather, schools should strive for a well-informed and educated staff on the attributes of group achievement tests.

Potential uses for achievement test data. Mehrens and Lehmann (1987) assert that "measurement and evaluation are essential for sound decision making" (p. 3). School professionals are required to make numerous decisions, or assist individual students, parents, and the public in making decisions. Good decisions are based on both the quality and quantity of relevant information (Mehrens & Lehmann, 1987). Standardized achievement tests can be an important factor facilitating sound decision making.

The Committee to Develop Standards for Educational and

Psychological Testing (1985) reported the following:

> [Achievement] test results are used by school administrators, teachers, parents, students, various citizen groups, and the media. The results of carefully selected, appropriate tests, when interpreted properly, can provide administrators with pertinent information about the general academic development and level of functioning of individual students, thereby helping to provide students with appropriate instruction and resources. Test scores can help teachers, students, and parents identify the specific academic strengths of a student on which to build, as well as the specific less-developed areas in need of remediation and special attention. (p. 49)

Based on the periodic administration of group achievement tests, Anastasi (1982) notes two critical uses, namely that achievement tests facilitate learning and assist educators in the identification of instructional needs. For Fall testing programs, a distinct advantage is the opportunity to effectively adjust instruction to meet a student's or the entire class's academic level (Anastasi, 1982).

The two most common purposes reported by publishers of group achievement tests include the following:

(a) improving within-classroom decisions, such as

> helping to diagnose each student's strengths and weaknesses in an area, and (b) improving educational decision making that occur outside the classroom, such as evaluating programs, curricula, and schools. (Nitko, 1989, p. 31)

It is beyond the scope of this paper to address the latter issue, but it is discussed by Airasian and Madaus (1983), Haertel and Calfee (1983), Mehrens and Lehmann (1987), and Lyman (1991).

Stone, Cundick, and Swanson (1988) are the only researchers to investigate specifically a group achievement test as a screening measure for identifying students potentially in need of special education services. The investigators suggested that a cutoff score at the 5th percentile (nationally) would reasonably include any child in need of special education services. Although this percentile score included only one percent of students in regular education, it missed almost 50 percent of students receiving special education services.

On the other hand, a cutoff score at the 20th percentile included almost 90 percent of the students in special education and 10 percent of students in regular education (Stone et al., 1988). Although the researchers investigated the utility of a group achievement test for the purpose of identifying children in need of special

education, they failed to address students who were in need of other support programs. In other words, were any of the 10 percent of students in regular education that scored below the 20th percentile participating in support services other than special education? Additionally, is the national percentile score the best comparison when teachers are comparing a student's achievement against others within the classroom? Are local percentiles a more accurate comparison worthy of investigation?

Recommendations have been regularly reported regarding the usefulness of achievement test data. In practice, however, many school systems are merely compiling results for publication to satisfy community curiosity or enhance school/community relations. Unfortunately, this means that little or no further interpretation is done, and test scores are simply deposited in file folders or recorded on filing cards (Mitchell, 1992).

Review of popular group achievement tests. The four most widely used and highly regarded group-administered achievement testing batteries include the Metropolitan Achievement Test-Seventh Edition (MAT7), California Achievement Test (CAT), Standford Achievement Test (SAT), and the Iowa Test of Basic Skills (ITBS) (Mitchell, 1992). A review of these achievement tests by Salvia and Ysseldyke (1995) considered the MAT and SAT as adequate instruments

for screening purposes based on test development, standardization, reliability, and validity. Due to internal-consistency variability, they caution users of the ITBS to check the manual for the reliability of specific grades and subtests. Additionally, there are no data on the 1993 edition to support either the construct validity or criterion-related validity. Regarding the CAT, information on standardization, reliability, and validity is either incomplete or limited.

Supportive Programs Beyond the Classroom

Most school systems have several available resource options for students with academic difficulties. Interventions for students who are academically at-risk usually vary according to the purpose of the program or the service delivery (Witt, et al., 1994). Direct services to the student may be provided by either placement in structured programs, indirect services provided by the classroom teacher consulting with a program manager, or a combination.

Students experiencing reading and/or math difficulties may receive additional instruction provided by programs funded federally by the Chapter 1 Program (U. S. Department of Education, 1993). The purpose of Chapter 1 is to:

> improve the educational opportunities of educationally deprived children in high poverty schools by helping

those children--

- succeed in the regular classroom
- attain grade-level proficiency [and]
- improve achievement in basic and more advanced skills.

These purposes are accomplished through supplemental education programs, schoolwide programs, and increased involvement of parents in their children's education. (p. 1)

Due to the socioeconomic criteria for Chapter 1 funding, school districts serving higher income level families may choose to develop a similar program funded at the local level. Similar objectives may be retained, but other criteria, such as teacher responsibilities, may be specifically tailored to match the academic needs of the student population. Further, districts may differentiate the provision of services at a building level, depending on student needs. Programs between buildings may range from structured pullout programs to services that are solely consultative.

In response to the high rate of student referrals for special education services, problem-solving teams have been either initiated by individual districts or state mandated (e.g., Nebraska Department of Education, 1992) (Witt, et al., 1994). Team composition is intended to be dependent on

the concerns regarding individual students. The Student Assistance Team (SAT) is a type of problem-solving team to brainstorm strategies that can be implemented by the classroom teacher in an attempt to address the student's academic and/or behavioral difficulties within regular education.

Another problem-solving team approach is the Individual Assistance Team (IAT) whose purpose is similar to the SAT process, but the service delivery procedures differ (Roosa & Biller, 1995). For example, SATs usually meet as a group while the IAT model establishes a one-to-one collaborative perspective. Additionally, measurable behavioral objectives, the method of data collection, and the evaluation procedure are implicitly established prior to the implementation of any interventions (for a full discussion, see Roosa & Biller, 1995).

Students continuing to demonstrate academic difficulties may be referred by the SAT to the Multidisciplinary Team (MDT) for consideration for a special education assessment (e.g., Nebraska Department of Education, 1992). Individual states establish specific criteria based on federal guidelines, and the assessment data determine whether a student qualifies for special education services within that state.

Purpose of the Study

The purpose of this study is to determine the usefulness of scores derived from a group achievement test as predictors for additional academic support provided by remedial programs. Thus, the primary research question becomes:

> Does a student's performance on the composite scores and/or the total score of a group achievement test, reliably discriminate between participation and nonparticipation in a support program beyond the classroom?

It seems that teachers tend to compare a student's rate of learning to the rates of peers in the class or the class as a whole. Therefore, local norms will be investigated in addition to national norms to test the hypothesis that local norms would most likely resemble teacher comparisons.

A stringent criterion of one standard deviation below the mean is often used to distinguish students performing within the average range from those students performing within the below average range. By definition, one standard deviation differentiates scores that are deviating significantly from the norm (Anastasi, 1982). Erroneous factors (i.e., motivation level, anxiety, illness) which contribute to an inaccurate estimate of academic performance, but are unlikely to affect significantly (i.e.,

one standard deviation) a student's total performance, also need to be considered. Based on this rational, the use of one standard deviation below the mean as a viable cutoff score for identifying students in need of further assistance will be investigated. Since most standardized tests have a mean of 100 and standard deviation of 15, the 50th percentile corresponds with the mean and the 16th percentile corresponds with one standard deviation below the mean. Therefore, the 16th percentile will be used to examine potential cutoff scores.

Three additional research questions are raised from the above discussion:

1. Will students performing one standard deviation or more below the mean in either two or three areas (e.g., combinations of the composite scores and the total score) significantly increase the likelihood of participation or nonparticipation in a support program beyond the classroom?
2. For those students participating in a support program, will different percentile levels at or below the 16th percentile significantly differentiate participation in a specific support program?

3. Will a student's performance on the different test scores (e.g., composite scores and the total score) increase the likelihood of a specific special education verification?

The answers to the above questions will contribute to this growing body of research, which has been primarily concerned with the early identification of students who are experiencing academic difficulties in the classroom. As Ogden and Germinario (1988) point out, the identification of students who are academically at-risk is the first step in increasing the effectiveness of our schools. Based on this initial information, educators can then monitor a student's performance within the classroom. Interventions could then be efficiently implemented which target the difficulties identified by the group achievement test, or the group achievement test data could be discounted as the scores are not reflective of the student's daily performance.

Chapter 2

Method

Participants

A total of 2,382 students from 27 elementary schools in a large Midwestern, metropolitan school district participated in this investigation. The average class size was 21.6, and only 5% of the total student population received free or reduced lunches. In this school district, the first opportunity to screen an entire population of students via a standardized group achievement test was at the third and fourth grades. During the 1994-95 school year, kindergarten screens were not conducted and the group achievement test was only administered to two grades at each building level (i.e., elementary, middle, high school). Participants consisted of 1,235 third grade students (626 boys, 609 girls) and 1,147 fourth grade students (578 boys, 569 girls). Students who had been previously identified for special education services were the only exclusions from participation, as this study investigated the initial identification of students who were academically at-risk, not those students who had already been identified.

Materials

The Metropolitan Achievement Test-Seventh Edition (MAT7) is a standardized, norm-referenced group achievement test (Psychological Corporation, 1994). It is divided into

14 levels, spanning grades K through 12. Acceptable norming procedures, reliability, concurrent validity, and construct validity are well documented in the technical and norming manuals (Finley, 1995). The publisher (Psychological Corporation, 1994) notes that content validity evidence is best determined by examining the match between the Instructional Objectives listed in the manual and the district's or state's curricula.

The Basic Battery is intended to provide information about a student's performance "for establishing instructional priorities and grouping for instruction" (Finley, 1995, p. 603). Derived scores provided by the MAT7 include scaled scores, national and local percentile ranks, stanines, grade equivalents, normal curve equivalents, and functional reading levels at the elementary levels. Scores are reported for individual subtests, composite areas (i.e., Total Reading, Total Mathematics, Language), and total battery (i.e., Basic Battery).

The MAT7 was administered to the participant population during the publisher's recommended Fall testing period in 1994. The third grade students completed the Primary 2 Level, Form S, and the fourth grade students completed the Elementary 1 Level, Form S. Both levels were commensurate

with the Psychological Corporation's recommended testing levels for each population. Answer sheets were computer-scored by the Psychological Corporation.

Independent and Dependent Variables

The independent or predictor variables in this investigation are continuous variables consisting of the MAT7 Total Reading, Total Math, and Basic Battery percentile scores at the National and Local levels (Psychological Corporation, 1994). The dependent or criterion variable in this study is a dichotomous variable (i.e., yes/no) that represented group membership defined as either participating in at least one support program or not participating in any of the selected support programs.

A total of 433 participants received some form of either direct or indirect assistance by at least one of the support programs. Although some of the participants received services provided by more than one program, each participant was coded for analysis according to either participating or not participating in a support program, rather than summing the total number of participants in each of the separate supportive services.

The existing support programs within this school district that were commensurate across the elementary schools were selected for this investigation and grouped into four categories according to their purpose. Table 1

presents the support programs and the number of participants who received services.

Table 1

Number of Participants in Each of the Support Programs

Support Services	Participation (n = 433)*	Nonparticipation (n = 1,949)**
Chapter 1/READ	276	2,106
SAT/IAT	263	2,119
MDT	167	2,215
Special Education	89	2,293

Note. n = Number of Participants.

*Total number of participants who received services provided by a support program.

**Total number of participants who did not receive assistance from any of the supportive services.

READ = Reading, Enrichment, and Development; SAT = Student Assistance Team; IAT = Individual Assistance Team; MDT = Multidisciplinary Team.

The Chapter 1 and the READ (Reading, Enrichment, And Development) programs were grouped together as they targeted specific and direct reading intervention. The second grouping included the Student Assistance Team (SAT) and the

Intervention Assistance Team (IAT). Their purposes were to problem-solve interventions targeting either academic or behavior concerns. The interventions were generally indirect services implemented by the classroom teacher. The third group was the Multidisciplinary Team (MDT). The purpose of the MDT was to determine if further individual assessment was warranted for special education services. Lastly, the category of special education verification included participants who met federal and state guidelines for special education services during the 1994-95 school year.

Table 2 presents the total number of participants within each of the following special education categories: Specific Learning Disability (SLD), Speech/Language Impairment (SLI), SLD/SLI, Behavior Disorder (BD), SLD/BD, Visual Impairment, and Other Health Impairment.

Table 2

Number of Participants in Each of the Special Education Verifications

Special Education Verification	Total Number of Participants
Specific Learning Disability (SLD)	41
Speech/Language Impairment (SLI)	10
SLD/SLI	28
Behavior Disorder (BD)	4
SLD/BD	3
Visual Impairment	2
Other Health Impairment	1

Note. Numbers represent the total number of participants for each verification category within special education.

Procedure

Fourteen school psychologists employed by the school district collected the data from each of their assigned elementary schools. A secretary within the Psychological Services Department prepared a list of participants who met the selection criteria (i.e., exclusion of students who qualified for special education services at the time of test administration) for each elementary school.

Coding forms were provided for recording the data, and included the following columns and coding procedures: name;

participant number; grade; gender; the local and national percentile scores on Total Reading, Total Math, and Basic Battery; categorical scoring (yes = 1; no = 2) for READ/Chapter 1, SAT/IAT, MDT, and SPED (Special Education) programs; and SPED classification. The name column was used for the sole purpose of facilitating ease of recording and accuracy for the school psychologists during the data collection. To ensure confidentiality, the name column was removed from the coding forms before the investigator's collection of each sheet.

Analyses

In order to investigate the degree of accuracy of the MAT7 percentile scores as reliable predictors for participation in a support program, a series of discriminant function analyses were generated (Tabachnick & Fidell, 1989). It should be noted a discriminant function analysis (DISCRIM) is the preferred method of analysis, rather than a regression analysis due to DISCRIM's classification procedures, DISCRIM's predictive ability, and the use of a dichotomous variable as the dependent variable.

Separate discriminant function analyses were run for Local percentile scores and National percentile scores. The discriminant functional analyses determined whether the two groups (participation or nonparticipation in a support program) could be reliably separated on the basis of the

predictor variables (MAT7 Total Reading, Total Math, and Basic Battery percentile scores).

The discriminant function analyses were performed by the SPSS-X (SPSS Inc, 1990) computer program. Prior to any discriminant function analyses, limitations of this analysis were addressed. First, because a one-way analysis was run, unequal sample sizes posed no special problem (Tabachnick & Fidell, 1989). The large number of cases in this study ensured robustness and avoided any difficulties of multicollinearity, singularity, and multivariate normality. Tests of tolerance, however, were also performed for each predictor, and predictors with insufficient tolerance were excluded. Finally, because discriminant function analysis is highly sensitive to inclusion of outliers, separate tests of univariate and multivariate outliers were run for the Participation and Nonparticipation groups. The results of these analyses did not warrant either the transformation or the elimination of significant outliers (i.e., severely skewed cases).

Given only two groups, there was one possibility of a statistically reliable discriminant function that separated the groups (Bordens & Abbott, 1988). The first goal was to interpret the degree of relationship between group membership and the set of predictors. A canonical correlation indicated the proportion of variance shared

between groups and predictors on the discriminant function. A correlation on the loading matrix in excess of .30 (9% of variance) between predictors and the function was considered eligible for interpretation (Tabachnick & Fidell, 1989).

The second goal was to evaluate the predictive value of the MAT7 scores by how well they separated the participants into the Participation and Nonparticipation Groups. The univariate *F* tested the reliability of the mean difference between the groups and showed how important a predictor was, by itself, in separating the members of each group (Tabachnick & Fidell, 1989). To avoid overinterpretation, only predictors with *F* ratios that were statistically significant after adjusting the error for the number of predictors in each set were considered (Tabachnick & Fidell, 1989). Adjustment was made on the basis of a Type I error rate for evaluating the contribution of the three predictors to between group contrasts. It was assumed that the predictors that met this criterion contributed a reliable, unique variance for separation of the two groups (Tabachnick & Fidell, 1989).

Once the predictors were found to discriminate significantly between group membership, a series of chi-square tests was performed to investigate the relationship between group membership (participation or nonparticipation in a support program) and combinations of MAT7 scores.

Cases were selected for analyses if the MAT7 scores were one standard deviation or more below the mean for either two MAT7 scores or for all three MAT7 scores. Separate chi-square tests were run for Local percentile scores and National percentile scores.

Each chi-square test compared the observed cell frequencies with the expected cell frequencies to determine the relationship between the two variables. The investigated cells included the following: (1) participants who had two scores that were one standard deviation or more below the mean and had participated in a support program; (2) participants who had all three scores that were one standard deviation or more below the mean and participated in a support program; (3) participants who had two scores that were one standard deviation or more below the mean and had not participated in a support program; and (4) participants who had all three scores that were one standard deviation or more below the mean and had not participated in a support program.

The results on the chi-square tests performed on these data showed whether a significant relationship was present between a combination of scores and whether participants were or were not in a support program.

Since group membership was unequal, a series of proportional analyses was generated to investigate the

relationship between students' scores and group membership within a specific support program. The following percentile ranges were selected so the results could be generalized beyond this population sample and compared with previous research. Table 3 presents the four percentile ranges which were used to evaluate potential cutoff scores for each of the MAT7 sections. The ranges coincide with most standardized testing instruments whose standard score mean is 100 and standard deviation is 15.

Table 3

Percentile Ranges Used to Investigate Cutoff Scores for the Support Groups by Each of the MAT7 sections

Range	Standard Deviation Ranges (Below the Mean)		Percentile Scores Ranges
A	1/2 SD to 1 SD	=	30th to 16th
B	1 SD to 1 1/2 SDs	=	15th to 7th
C	1 1/2 SDs to 2 SDs	=	6th to 2nd
D	> 2 SDs	=	1st

Note. SD = standard deviation.

The percentile scores in Range A represented participants who were directly adjacent to the one standard

deviation (16th percentile) cutoff level. This range enabled comparisons to determine if significant differences existed between percentile groupings that were within proximity of one standard deviation (16th percentile) from the mean. Each of the ranges on the MAT7 sections was tested for significance of difference between group membership (participation or nonparticipation in each of the support programs) at the .05 level (± 1.96). An example of this analysis would be as follows: All participants who scored from the 16th percentile to the 30th percentile (Range A) on the MAT7 Total Reading composite and participated in the Chapter 1/READ category were compared to all students who scored from the 16th percentile to the 30th percentile (Range A) on the MAT7 Total Reading composite and had not participated in the Chapter 1/READ category. This analysis was repeated for each percentile range by each support program category. The results determined whether a significant difference existed between group membership at differing percentile ranges.

Lastly, a series of proportional analyses was conducted to determine if Local and National percentile scores differentiated participants within the specific special education verifications. Groups included the following: Specific Learning Disability (SLD), Speech/Language Impairment(SLI), and SLD/SLI. Participants who were

verified with a Behavior Disorder, SLD/BD, Vision Impairment, and Other Health Impairment were excluded from analysis; the small number of participants within each of these categories would have yielded results that overestimated the differences as significant whenever a specific range had zero number of cases. An example of this analysis would be participants who were verified as SLD and scored between the 16th percentile to the 30th percentile (Range A) on the MAT7 Total Reading composite and were then compared to participants who qualified for special education services, but were not verified as SLD and scored within Range A on the same composite.

Chapter 3

Results

Table 4 displays the means and standard deviations of the Metropolitan Achievement Test-Seventh Edition (MAT7) percentile scores for the Reading Composites, Math Composites, and the Basic Batteries at the Local and National levels for this sample population.

Table 4

Means and Standard Deviations by MAT7 Sections

MAT7 Sections	Mean	Standard Deviation
Local Reading	50.5	28.7
Local Math	49.9	28.9
Local Basic Battery	50.3	28.8
National Reading	64.6	25.5
National Math	66.5	25.8
National Basic Battery	66.5	25.1

Note. The means and standard deviations are derived from the MAT7 percentile scores.

The local scores represent an even distribution of scores with overall means at approximately the 50th percentile. The mean for the participants' scores at the National level was significantly higher ($p < .01$, t test) on

the MAT7 (Average Mean = 65.9), demonstrating a consistently superior performance by this population sample compared to the national norm group. Standard deviations at the National level are smaller than at the Local level, suggesting greater measurement variability within the district (Local level).

Table 5 displays the percentage of participants by gender and grade who were (a) one standard deviation or more below the mean, (b) within one standard deviation of the mean, or (c) one standard deviation or more above the mean for each of the MAT7 sections.

Table 5

Percentage Distribution of Scores for the MAT7 Sections by Grade, Gender, and Total Population

MAT7		Below	Average	Above
Local Reading	Male	24	56	20
	Female	19	61	20
	Grade 3	22	59	19
	Grade 4	21	59	20
	Total	22	58	20
Local Math	Male	19	57	24
	Female	24	56	20
	Grade 3	22	56	22
	Grade 4	21	57	22
	Total	22	56	22
Local Basic Battery	Male	23	57	20
	Female	21	58	22
	Grade 3	22	57	21
	Grade 4	21	57	21
	Total	22	57	21
National Reading	Male	22	57	20
	Female	18	62	20
	Grade 3	20	63	17
	Grade 4	20	57	23
	Total	20	60	20

Table 5 (cont.)

MAT7		Below	Average	High
National Math	Male	18	64	18
	Female	22	63	14
	Grade 3	22	61	17
	Grade 4	18	67	15
	Total	20	64	16
National Basic Battery	Male	20	64	16
	Female	19	64	17
	Grade 3	21	64	15
	Grade 4	18	65	18
	Total	20	64	16

Note. Numbers represent percentages. Numbers are rounded to the nearest whole number and, therefore, may not equal 100.
Below = percentage of participants whose scores are one standard deviation or more below the mean; Average = percentage of participants whose scores are within one standard deviation of the mean; Above = percentage of participants whose scores are one standard deviation or more above the mean.

The distribution of scores across participants for each of the MAT7 sections (Local and National levels) suggests a

normal bell-shaped curve, as most of the participants are within the average range and fewer participants are within the below or above average ranges. In addition, the total percentages of participants scoring within the below or above average ranges are fairly evenly distributed between these ranges for all the MAT7 sections.

In the area of reading at the National and Local levels, a higher percentage of males scored below the average range, while the females had a higher percentage within the normal range. The percentage of males and females scoring above the average range in the area of reading was similar. Male participants had more difficulty on the reading section of the MAT7 compared to the females.

Grade differences are noted at the National level in reading, with a higher percentage of participants in Grade 3 scoring within the average range in contrast to the above and below average ranges. A higher percentage of participants in Grade 4 than in Grade 3 scored above the average range. Percentages at the Local level were similar across grade levels.

A larger percentage of females scored below the average range in the area of math at the National and Local levels, while a greater percentage of males scored within the above average range. Little difference is noted within the average range for gender. Overall, on the math section of

the MAT7 at the National and Local levels, males had greater success indicated by fewer males scoring below the average range and more males scoring above the average range.

Comparing grade differences in the area of math at the National and Local levels, the participants in Grade 4 had fewer scores within the below average range and a higher percentage of scores that fell within the average range at the National level. However, the participants in Grade 3 had a higher percentage of scores within the above average range. Overall, the participants in Grade 4 had a slightly higher rate of success on the MAT7 compared to Grade 3 at the National level (a difference of 4 percent).

Comparing the basic battery scores, gender differences existed at the Local level, with fewer females scoring below the average range and a higher percentage of females scoring above the average range. A higher percentage of participants in Grade 3 scored within the below average range, while a higher percentage of participants in Grade 4 scored above the average range. Overall, females and participants in Grade 4 had the greatest level of success on the total MAT7 battery.

Comparisons across the MAT7 sections at the National level yielded a higher percentage of participants scoring within the above average range in the area of reading. Percentage distributions at the Local level revealed more

participants scoring within the above average range and fewer participants falling within the average range in the area of math. Lastly, a higher percentage of participants scored outside the average range at the Local level than at the National level, which is consistent with the smaller standard deviations at the National level.

Question #1: Does a student's performance on the composite scores or the total score of a group achievement test significantly discriminate between participation or nonparticipation in a support program beyond the classroom?

This question was addressed by evaluating the degree of classification accuracy generated from a series of discriminant function analyses using the MAT7 percentile scores at the National and Local levels. The Reading Composite, Math Composite, and Basic Battery were the predictor (independent) variables and group membership (participation and nonparticipation in a support program) was the criterion (dependent) variable.

The Participation Group consisted of participants who had received some form of assistance either directly or indirectly from the Read and/or Chapter 1 programs, the

Student Assistance Team (SAT) and/or Individual Assistance Team (IAT), the Multidisciplinary Team (MDT), and/or Special Education. Cases ranged from participants who were in only one of the supportive services (i.e., Read/Chapter 1, SAT) to participants who had received assistance from all the support programs. The Nonparticipation Group consisted of participants who did not receive any form of assistance provided by the support programs. Categorization for the Participation and Nonparticipation groups was, therefore, based on students having either received or not received some form of service.

Separate discrimanant function analyses using the SPSS-X (SPSS Inc., 1990) computer program were executed for the Local and National percentile scores. All discriminant analyses specified a minimum tolerance level of .001, which was calculated from the determinant of the within-cell correlation matrix to detect singularity and multicollinearity. All predictor variables passed the minimum tolerance level test at the .001 level.

Local Percentile Scores. Table 6 presents the group means and standard deviations for participants in a support program and participants not in a support program in the investigation.

Table 6

Group Means and Standard Deviations of the Local Percentile Scores by Group Membership

Group Membership		Local		
		Reading	Math	Basic Battery
Participants in a Support Program (n=433)	Mean	24.7	29.0	23.3
	SD	24.1	25.9	23.3
Participants not in a Support Program (n=1,949)	Mean	56.3	54.5	56.3
	SD	26.3	27.5	26.4

Note. n = Number of participants.

A direct discriminant function analysis using the Local Reading Composite, Math Composite, and Basic Battery percentile scores of the MAT7 as the predictor variables and group membership as the criterion variable revealed a significant discrimination between the two groups [$X^2(3, N = 2,382) = 522.52$, $p < .0001$], whereas 3 represents the degrees of freedom, N is the number of participants, and 522.52 is the chi-square calculation.

The canonical correlation between the three predictor variables and group membership was .44, accounting for 20% of the variance (lambda = .80). Lambda is the proportion of

variance of the dependant variable that is not accounted for by the independent variables (error variance).

The standardized discriminant equation for Local Reading (X^1), Local Math (X^2), and Local Basic Battery (X^3) derived from this procedure is as follows:

$$D = 0.12670(X^1) - 0.14158(X^2) + 0.99079(X^3)$$

where D is the discriminant function score calculated for each participant, 0.12670 is the standardized raw score for Local Reading, -0.1458 is the standardized raw score for Local Math, and 0.99079 is the standardized raw score for Local Basic Battery. Since SPSS-X, by default, standardizes the discriminant function coefficients the mean of D becomes zero and the standard deviation of D is 1 (Marascuilo & Levin, 1983).

The group centroid for the Participation Group (Group 1) was -1.05116, whereas, the group centroid for the Nonparticipation Group (Group 2) was 0.23353. The centroid represents the mean of the discriminant function scores within each group (Tabachnick & Fidell, 1989).

The discriminant function scores were used to classify the participants into either the Participation (Group 1) or Nonparticipation Groups (Group 2). If a D score was below zero, the case was classified into Group 1. If a D score

was above zero, the case was classified into Group 2. The greater the distance the participant's discriminant function score was from the mean of zero, the greater the likelihood that the participant was classified into the correct criterion group (Marascuilo & Levin, 1983).

Table 7 shows the classification results in which 80% of the Participation Group and 69% of the Nonparticipation Group were accurately classified. These results yielded an overall classification accuracy of 71 percent.

Table 7

Classification Accuracy of the Local Percentile Scores

Actual Group	Number of Cases	Predicted Group Membership	
		Group 1	Group 2
Participation in a Support Program (Group 1)	433	345 79.7%	88 20.3%
Nonparticipation in a Support Program (Group 2)	1,949	596 30.6%	1,353 69.4%

Note. The hit rate for the Participation Group is in the Group 1 column and the miss rate is in the Group 2 column. The hit rate for the Nonparticipation Group is in the Group 2 column and the miss rate is in the Group 1 column.

For participants in a support program (Participation Group, n = 433), 345 cases were accurately classified and 88 cases were misclassified. Based on the disciminant function analysis, 88 participants did not receive any form of additional assistance beyond the classroom although their performance on the MAT7 suggested a need compared to their peers. For participants not in a support program (Nonparticipation Group, n = 1,949), 1,353 cases were classified correctly and 596 cases were misclassified. Based on the participants' performances on the MAT7, 596 students received additional support, although their performance on the MAT7 did not suggest a need for additional supportive services.

National Percentile Scores. Table 8 displays the group means and standard deviations of the National percentile scores.

Table 8

Group Means and Standard Deviations of the National Percentile Scores by Group Membership

Group Membership		National		
		Reading	Math	Basic Battery
Participants in a Support Program (n = 433)	Mean	39.2	45.9	40.2
	SD	25.2	27.7	24.4
Participants not in a Support Program (n = 1,949)	Mean	70.3	71.1	72.3
	SD	21.9	23.0	21.3

Note. n = Number of participants.

Although the mean values at the National level were consistently higher than the Local mean values, the difference between the means of the criterion groups was consistent (i.e., Local Reading, 56.3 - 24.7 = 31.6; National Reading, 70.3 - 39.2 = 31.1). Contrasting the standard deviations at the Local and National levels revealed smaller SDs at the Local level for the Participation Group; however, the Nonparticipation Group at the National level had considerably smaller SDs.

The second direct discriminant function analysis using the National Reading Composite, Math Composite, and Basic Battery percentile scores as predictor variables and group

membership as the criterion variable also showed a significant separation of the groups [$X^2(3, N = 2,382) = 672.55, p < .001$]. The canonical correlation of .50 between the three MAT7 percentile scores and group membership accounted for 25% of the variance (lambda = .75). The National percentile scores accounted for an additional 5 percent of variance compared to the Local percentile scores.

The standardized discriminant function derived from the National percentile scores of the MAT7 Reading Composite (X^1), Math Composite (X^2), and Basic Battery (X^3) is as follows:

$$D = -0.07718(X^1) - 0.26015(X^2) + 1.26948(X^3)$$

where D is the discriminant function score calculated for each participant, -0.07718 is the standardized raw score for National Reading, -0.26015 is the standardized raw score for National Math, and 1.26948 is the standardized raw score for National Basic Battery.

The group centroid for the Participation Group (Group 1) was -1.21231, whereas the group centroid for the Nonparticipation Group (Group 2) was 0.26933. The centroid at the National level for the Participation group was at a greater distance (0.16) from the mean of zero than the Local

percentile scores. Group centroids for the Nonparticipation Group at the Local and National levels were similar.

The discriminant function scores derived from the National percentiles were also used for each case to classify the participants into either the Participation (Group 1) or the Nonparticipation (Group 2) Groups. Table 9 displays the National percentile scores of the MAT7 in which 77% of the Participation Group and 78% of the Nonparticipation Group were accurately classified. These results yielded an overall classification accuracy of 77.5 percent.

Table 9

Classification Accuracy of the National Percentile Scores

Actual Group	Number of Cases	Predicted Group Membership	
		Group 1	Group 2
Participation in a Support Program (Group 1)	433	334 77.1%	99 22.9%
Nonparticipation in a Support Program (Group 2)	1,949	438 22.5%	1,511 77.5%

Note. The hit rate for the Participation Group is in the Group 1 column and the miss rate is in the Group 2 column. The hit rate for the Nonparticipation Group is in the Group 2 column and the miss rate is in the Group 1 column.

The hit rate for the Nonparticipation Group was considerably higher for the National percentile scores (i.e., National = 78%; Local = 69%). However, the hit rate for the Support Group was higher for the Local percentile scores (i.e., Local = 80%; National = 77%).

Question #2: Will students performing one standard deviation or more below the mean in either two or three areas (e.g., a combination of the composite scores and the total battery score) significantly increase the likelihood of participation or nonparticipation in a support program?

Separate chi-square tests for the National and Local percentile scores were generated by the SPSS-X (SPSS Inc., 1990) computer program to assess the relationship between combinations of percentile scores and the Participation and Nonparticipation Groups. Cases were classified into: (a) Group 1 if only two MAT7 scores were at or below one standard deviation from the mean; or (b) Group 2 if all three of the MAT7 scores fell at or below one standard deviation from the mean. Groups 1 and 2 were mutually exclusive (i.e., participants meeting criteria for inclusion into Group 2 were not classified into Group 1). Participants with no scores or only one score that fell one standard deviation or more below the mean were excluded from the chi-square analyses.

Local Percentile Scores. Cases were selected if percentile scores were at or below the 21st percentile, as the mean of the percentile scores at the Local level was 50

and the standard deviation was 29. Table 10 displays the results of the contingency table.

Table 10

Contingency Table for Local Percentile Scores

MAT7 Combinations	Participation in a Support Program	Nonparticipation in a Support Program
Two Scores	81	123
Three Scores	175	107

Note. Two Scores = Cases with two scores of the MAT7 Local Reading, Math, or Basic Battery that were 1 SD or more below the mean; Three Scores = Cases with all three of the MAT7 scores 1 SD or more below the mean.

The chi square test revealed a significant relationship between group membership, and whether participants performed at least one standard deviation or more below the mean ($P \leq 21$) on either two or all three of the Local MAT7 scores [$X^2(1, N = 486) = 23.72, p < .001$], whereas 1 equals the degrees of freedom, N is the number of participants, and 23.72 is the chi-square calculation. Participants scoring at or below the 21st percentile on all three measures were significantly more likely to participate in a support group

($p < .01$, binomial test). In contrast, participants who scored at or below the 21st percentile on two measures were significantly more likely to not participate in a support group ($p < .01$, binomial test).

National Percentile Scores. Cases were selected if percentile scores were at or below the 40th percentile as the mean of the percentile scores at the National level was 66 and the standard deviation was 26. Table 11 displays the results of the contingency table.

Table 11

Contingency Table for National Percentile Scores

MAT7 Combinations	Participation in a Support Program	Nonparticipation in a Support Program
Two Scores	81	107
Three Scores	162	88

Note. Two Scores = Cases with two scores of the MAT7 National Reading, Math, or Basic Battery that were 1 SD or more below the mean; Three Scores = Cases with all three of the MAT7 scores 1 SD or more below the mean.

The chi-square test for the National scores indicated a significant relationship between group membership and

whether students performed at least one standard deviation or more below the mean ($P \leq 40$) on either two or all three of the national scores [$X^2(1, N = 438) = 20.49, p < .01$]. Participants scoring at or below the 40th percentile on all three measures were significantly more likely to participate in a support group ($p < .01$, binomial test). These results were expected, due to the high correlations that are naturally inherent when the same scores are compared between different norm groups (e.g., Basic Battery scores at the National and Local levels). This finding is substantiated by the proportional increases in the means for each set of scores for this population sample (refer to Table 1).

Question #3: For those students participating in a support program, will different percentile levels that are one standard deviation or more below the mean significantly differentiate participation in a specific support program?

A series of proportional analyses was conducted for each of the MAT7 percentile scores at the Local and National levels for Total Reading, Total Math, and Basic Battery by

each of the support programs (Read/Chapter 1, SAT, MDT, and Special Education). The formula for the difference between two proportions is

$$z = \frac{P_1 + P_2}{\sqrt{\frac{P_1(1 - P_1) + P_2(1 - P_2)}{N_1 + N_2}}}$$

where P_1 = the proportion of Group 1, P_2 = the proportion of Group 2, N_1 = number of participants in Group 1, and N_2 = number of participants in Group 2 (Bruning & Kintz, 1968). A significant difference existed if z was at least $\pm$ 1.96 ($p \leq .05$).

The following percentile ranges were selected for analysis as they coincide with most standardized testing instruments that have a standard score mean of 100 and standard deviation of 15. Using the percentile scores that coincide with the standard deviations within this population sample would have overinflated the results due to the proportional increments between the National and Local scores, as well as the inherent correlations between the

National and Local scores. The results of the proportional analyses could then be generalized beyond this population sample and compared with other research. Based on this justification, the 16th percentile becomes the cutoff score as it coincides with one standard deviation below the mean.

Table 3 presents the percentile ranges for Group 1 and consisted of participants who scored between the: (a) 16th to 30th percentiles (Range A); (b) 7th to 15th percentiles (Range B); (c) 2nd to 6th percentiles (Range C); or (d) 1st percentile (Range D) on the MAT7 sections and received services provided by a specific support program. Inclusion of Range D (16th through the 30th percentile) allowed for group comparisons within proximity of one standard deviation below the mean. Participants who were not in a support program but scored within each of the above MAT7 score ranges were classified into Group 2. The N for Group 1 was the total number of participants scoring at or below the 30th percentile who were in the specific support program. The N for Group 2 was the total number of participants whose score was at or below the 30th percentile and not in the support program.

Table 12 displays the proportional analyses calculations. Cutoff scores below the 16th percentile were used when significant differences were found in favor of supportive services for Ranges B, C, and D, as well as

significant differences in favor of no supportive services for participants scoring within Range A.

Cutoff scores below the 7th percentile were used when Ranges C and D significantly differentiated participants who were in a support program from participants who were not in a support program, Range B was not significant, and Range A was significant for participants who were not in a support program.

Table 12

Proportional Analyses

Percentile Ranges	16th Percentile		7th percentile	
	Support	No Support	Support	No Support
A (16th-30th)		**		**
B (7th-15th)	**		ns	ns
C (2nd-6th)	**		**	
D (1st)	**		**	

Note. ** = z scores that were at least $\pm$ 1.96 ($p < .05$); ns = nonsignificant.

Table 13 displays the results of the proportional analyses in which consistent results for significant differences were found for the different percentile ranges.

Table 13

Results of the Proportional Analyses for the Support Programs

MAT7 Sections	READ or Chapter 1	SAT or IAT	MDT	Special Education
Local Reading	<16th	<7th	<7th	<7th
Local Math	<16th	<7th	<7th	<7th
Local Basic Battery	<16th	<7th	<16th	<16th
Natl Reading			<16th	<16th
Natl Math		<16th	<7th	<7th
Natl Basic Battery				

Note. $p \leq .05$

<16th = Ranges B, C, and D were found to differentiate significantly group membership for participants who were in a support program and Range A significantly differentiated group membership for participants who were not in a support program.

<7th = Ranges C and D were found to differentiate significantly group membership for participants who were in a support program, Range B did not significantly differentiate participants, and Range A significantly differentiated group membership for participants who were not in a support program.

Percentile scores that significantly differentiated ($p \leq .05$) between participants who were in a support program below the 16th percentile (Ranges B, C, & D) from participants who were not in a support program above the 16th percentile (Range A) were found for all the MAT7 sections except the National Basic Battery. However, using the 16th percentile score as the cutoff score varied across the MAT7 sections by support groups.

A cutoff score at the 16th percentile on the Local Reading, Local Math, and Local Basic Battery was found to differentiate significantly ($p \leq .05$) those students who participated in the Chapter 1 and READ programs from those students who had not participated in these supportive services.

For participants who indirectly received services provided by SAT or IAT, percentile scores below the 7th percentile on the Local Reading, Local Math, and Local Basic Battery significantly differentiated the provision of these services from participants who had not received either of these services. Additionally, scores below the 16th percentile on the National Math were found to differentiate significantly participation in the SAT and IAT from students who had not received these services.

Lastly, the results for MDT and Special Education services were identical for the cutoff scores and the

specific MAT7 sections. Scores below the 16th percentile on the Local Basic Battery and National Reading, as well as scores below the 7th percentile on the Local Reading, Local Math, and National Math, were found to differentiate significantly participants who were referred to MDT and qualified for Special Education services from participants who were not referred to MDT or qualified for Special Education services.

The remainder of the results comparing the proportional differences at the specified ranges for the MAT7 percentile scores with the different support programs yielded inconsistent findings. For example, the results of the National Reading scores for the READ/Chapter 1 support services category were insignificant for Ranges D (1st percentile) and C (2nd through the 6th percentiles), significant for the support services for Range B (7th through the 15th percentiles), and significant for nonparticipation in support programs for Range A (16th through the 30th percentiles).

Question #4: Will a student's specific performance on the different test scores (e.g., composite scores and the total score) increase the likelihood of a specific special education verification?

To answer this final question, a second series of proportional analyses was conducted for each of the MAT7 percentile scores at the Local and National level for Reading, Math, and Basic Battery by each of the following special education verifications: Specific Learning Disability (SLD), Speech/Language Impairment (SLI), and SLD/SLI.

The same formula and percentile ranges from question #3 were used to answer this question, in which Group 1 consisted of participants who were verified within a specific special education category, and scored between the: (a) 16th to 30th percentiles (Range A); (b) 7th to 15th percentiles (Range B); (c) 2nd to 6th percentiles (Range C); or (d) 1st percentile (Range D) on the MAT7 sections and was verified within a specific special education category. Group 2 encompassed participants who qualified for special education services other than the category under comparison, but scored within each of the above MAT7 score ranges. The N for Group 1 was the total number of participants scoring at or below the 30th percentile who were verified within the

specific special education category. The N for Group 2 was the total number of participants who scored at or below the 30th percentile, and who were verified within a special education category other than the category under analysis.

The results of the proportional analyses were completely inconsistent (Table 14). Although a substantial number of specific ranges within the different special education verification categories were found to differentiate significantly within a given range, cutoff scores were not found. For instance, when participants who were verified as SLD were compared with participants who had different verifications, significant differences were yielded in favor of an SLD verification for Range B; but for Ranges A and C, significant differences were found in favor of not being verified as SLD. Lastly, within Range D, significant differences were not found between participants. Based on these results, neither cutoff scores nor performance on any of the specific MAT7 sections significantly differentiated participants among the individual Special Education verifications.

Table 14

Results of the Proportional Analyses for the Special Education Verifications

MAT7 Sections	Special Education Verifications		
	SLD	SLI	SLD/SLI
Local Reading	ns	ns	ns
Local Math	ns	ns	ns
Local Basic Battery	ns	ns	ns
Natl Reading	ns	ns	ns
Natl Math	ns	ns	ns
Natl Basic Battery	ns	ns	ns

Note. ns = nonsignificant.

Chapter 4

Discussion

The major question in this investigation was whether a student's performance on the composite scores or the total score of a group achievement test significantly discriminated between his or her subsequent participation or nonparticipation in a support program beyond the classroom. The Metropolitan Achievement Test-Seventh Edition's (MAT7's) Total Reading, Total Math, and Basic Battery percentile scores were found to collectively reveal a significant discrimination between the Participation and Nonparticipation Groups at the National and Local levels. Percentile scores at the Local level accounted for 20 percent of the variance for group membership, while a separate analysis of the percentile scores at the National level accounted for 25 percent of the variance. Other factors, such as socioeconomic status, work-related behaviors within the classroom, and intellectual ability (Bay & Bryan, 1992; Lloyd, 1978; Payne & Payne, 1991), may also be contributing to the total percent of variance accounted for by differentiation of group membership.

Applying the discriminant function equations to classify cases into the Participation and Nonparticipation groups yielded overall classification accuracies of approximately 78 percent for the National percentile scores

and approximately 71 percent for the Local percentile scores. The percentage of accurate classifications from the Local percentile scores (hit rate = 80%) was higher for the Participation Group than the National percentile scores (hit rate = 77%). When other supportive services were considered, these hit rates exceeded the commonly cited hit rate for teacher referrals for special education services alone (73%) (Salvia & Ysseldyke, 1995). The higher hit rate for the Local percentile score may be due in part to students receiving many of the additional supportive services based on teacher referrals (i.e., SAT, MDT). In other words, the Local percentile scores may be a reflection of teachers basing their referrals on comparisons with peers within the classroom.

School professionals also need to consider the miss rate for the Nonparticipation Group when interpreting the achievement test data. The miss rates at the National (22%) and Local (31%) levels represent students who received additional supportive services although their performance on the achievement test did not deem it necessary. In other words, the student's performance on the MAT7 was not commensurate with other factors, such as work-related behaviors in the classroom or the tolerance level of the teacher (Salvia & Ysseldyke, 1995). Because it is much more difficult to perform well than to perform poorly, one should

consider that the data for students with high scores are more likely to reflect actual performances. Therefore, it can be concluded that the miss rate of the Nonparticipation Group may represent any combination of the students: (a) receiving informal interventions (i.e., peer tutoring, reteaching strategies); (b) demonstrating poor test taking skills (e.g., anxious); (c) being under consideration for services in the future; (d) underachieving within the classroom; and/or (e) having teachers with lower tolerance levels. In contrast, the miss rate for the Participation Group may reflect a combination of such factors as the students overachieving within the classroom and/or greater tolerance levels by the teachers.

Overall, the results of this investigation suggest that the Local Percentile scores for this school district are a slightly better predictor of students being identified as academically at-risk and in need of additional support beyond the classroom (hit rate = 80%). However, to generalize these results to other populations, the more conservative and larger norm group would dictate the use of the National percentile scores until there has been further research investigating the use of Local percentile scores. It is disappointing that these results are not directly comparable to past research, as prior methods of investigation have focused on correlational findings between

teacher identification of students who are academically at-risk and achievement tests rather than classification accuracies. However, in general, the results support previous findings confirming the poor performance on a group achievement test as a valid factor contributing to an at-risk status (Lloyd, 1978; Payne & Payne, 1991; Stone et al., 1988).

Having established the predictive value of a group achievement test, the second goal was to investigate the relationship between combinations of percentile scores (i.e., Total Reading, Total Math, Basic Battery) that were one standard deviation or more below the mean and group membership. The data suggested that all three percentile scores that were one standard deviation or more below the mean increased the likelihood that students were provided additional support services. Further, at the Local level, students with two scores that were one standard deviation or more below the mean significantly more likely not to participate in a support program.

A more detailed analysis found several cutoff scores from the composite scores and total battery to differentiate significantly students whose performance on the MAT7 was below a specified cutoff point. All the Local percentile scores were found to differentiate significantly group membership by each of the support programs, although the

cutoff scores varied. The results yielded by the Local percentile scores further support the proposition that teacher referrals are based on comparisons between students within the classroom.

A cutoff score at the 16th percentile for each of the Local scores was found for those students who received additional reading remediation provided by the Chapter 1 or READ programs. For the teacher referred programs (i.e., SAT, IAT, MDT) and special education, a stricter cutoff score at the 7th percentile for each of the Local scores, except the Basic Battery for MDT and special education, was found to differentiate significantly students. Further, the National Reading composite score was found to differentiate significantly students at the 16th percentile for MDT referrals and special education, while the National Math score was at the 7th percentile for the same programs. Lastly, the National Math score could be used at the 16th percentile to differentiate students who were referred to SAT or IAT. The practice of peer comparisons may account for most of the higher cutoff scores at the National level (16th versus 7th percentiles).

Of particular interest was the lack of consistency needed for statistical significance on the National Basic Battery which did not differentiate group membership in any of the support programs. These results do not support the

recommendation by Stone et al. (1988) that the total battery score, from a different group achievement test, Standford Achievement Test (SAT), should be used with a cutoff score at the 5th percentile. However, they specifically compared the performances of students in regular education with students receiving special education services. It is not clear from the available data whether students perform differently on different group achievement tests, or whether this in an artifact for this specific school district since the mean of the participants' performances (66th percentile) were above the national mean.

The failure of cutoff scores to differentiate students within the special education population was not expected. The overlapping characteristics and difficulties that distinguish this population of students may have compounded this effort. For example, students who are verified with a specific learning disability in reading and students who are verified as having a receptive language impairment may be experiencing reading comprehension difficulties that would have a similar impact on the total performance on the group achievement test.

Implications for Practitioners

The results of this study should be useful to educators as they strive for efficient methods of identifying students who are academically at-risk. Before educators can design a

screening strategy based on group achievement test data, they need to identify whether their school's mean performance on the achievement test differs from the National mean (50th percentile). The results of this study suggest that: (1) if the means are different, the Local percentile scores are more likely to parallel the occurrence of classroom teachers' comparisons between peers; and (2) if the means are not different, the National percentile scores should be used as they reflect individual comparisons based on a larger representative norm group.

Once the determination has been made regarding the use of either the National or Local percentile scores, individual cases can be further investigated. Computers can ease this burden by selecting cases in which the total reading, total math, and basic battery scores are one standard deviation or more below the mean. From the list of students who are potentially at-risk, further analysis of students who meet criteria by specific cutoff scores can be grouped according to the individual school system's support service options.

Given this initial information, program coordinators and psychologists can work directly with the classroom teachers to monitor the student's classroom performance and confirm or disconfirm the initial suspicions. This author's basic position is that a disconfirmatory approach needs to

dictate this step of the process. In so doing, professionals not only guard against the potentially negative influence of labeling and/or grouping students based on one performance, but also the effects of a self-fulfilling prophecy by either teachers or the students. Initial interventions within the classroom targeting the possible weakness(es) identified by the group achievement test may alleviate these difficulties before they have an opportunity to impact further the student's rate of success within the classroom. In other words, early remediation of a few difficulties is essential to abate the possibility of compounded deficiencies leading to a devastating outcome of academic failure.

It needs to be stressed that the only advantage of an early identification process for students who may be academically at-risk is to provide the opportunity of academic success for students within the classroom, not the process by which to populate the available service options beyond the classroom.

Limitations

The purpose of this investigation was to determine whether a student's performance on a group achievement test was predictive of support services beyond the classroom rather than the prominent practice of student placement based on teacher initiated referrals. Given this

information, the student's performance on a group achievement test could alert educators to the need for further investigation in to the reasons underlying a poor performance. However, most of the support programs used in this study were dependent on teacher referrals. At a minimum, the results supported the predictive validity of a group achievement test for additional support services beyond the classroom. It is unclear from the findings whether these results are simply a confirmation of teacher referrals, or whether the results support a group achievement test as an initial academic screen. In any event, educators can efficiently monitor students and initiate early interventions, potentially reducing the miss rate for those students deemed in need of additional support services (as based on the test data).

Second, this participant population varied from a random population sample in the following ways: (1) students already receiving special education services were excluded from participation; (2) the participant sample was restricted to third and fourth graders; (3) the mean scores at the National level were significantly different from the norm group (66th percentile versus 50th percentile), and the free/reduced lunches for this population were significantly less (only 5%). Inclusion of participants in the primary grades (k-2) is needed in future investigations to increase

the participant sample and to encompass students who have yet to be quantitatively (i.e., individualized assessments) identified for special education services. Additionally, Rumberger (1987) discussed measures of parental involvement or the educational levels of parents as viable options as measures of low socioeconomic status (SES), rather than the use of free/reduced lunches. These alternative measures may have been a preferred method for possibly substantiating this participant population as representative of a random population sample on SES. Finally, replication of this study is needed to investigate further whether the significant mean differences at the National level have impeded generalizability.

Past research (Mitman, 1985) has addressed the question of potential misconceptions or altered expectations of teachers based on results derived by tests. Thus, a third limitation of this study was the teachers' accessibility of the data from the MAT7. Questioning teachers would have been advantageous to ascertain: (1) teachers who may have used achievement test data to initiate referrals; (2) teachers whose initial concerns were governed purely by the student's performance in the classroom; (3) teachers who may have used the data as collaborative evidence; or (4) teachers who may not have used the test data at all.

Fourth, some program coordinators across the different

elementary schools may have used the MAT7 data to create initial rosters of students targeted for services provided by the Chapter 1 and READ programs. It is unclear whether this may have accounted for the reading composite and possibly the basic battery as significant predictors. This limitation would not have accounted for the math composite scores also being a significant predictor.

Lastly, the MAT7 was given to this population for the first time, which impeded the possibility of a cross-validation or longitudinal study investigating the stability or variability of scores for students from year to year. Of specific interest for future research would be the investigation of the stability/variability of students' performances on a group achievement test who are identified as academically at-risk.

Future Research

Of particular interest for future research are other characteristics that may be contributing to the total variance accounting for students being identified as academically at-risk. The growing body of research literature related to an at-risk status is dedicated to the early identification of these students to remedy this potentially inevitable plight (Lloyd, 1978). Although some special education referral literature has concentrated on characteristics distinguishing this specific population from

students who are merely identified as academically at-risk (Bay & Bryan, 1992), initially investigating the characteristics accounting for these combined populations may narrow the plethora of characteristics to a manageable number and a more realistic application by practitioners.

Secondly, future research needs to investigate students who were misclassified by the discriminant function equation. For the students in the Nonparticipation Group who received services although the test data suggested otherwise, studies need to discover not only the specific program(s) in which they received services, but also the length of time that they received the services. Were these students receiving services as they needed additional "catch-up" at the beginning of the school year, or did these students' rates of academic learning during the school year no longer match other peers or teacher expectations? To answer these questions, future research needs to target not only the time of year that the students received additional supportive services, but also the length of time these services were provided.

In contrast, what factors disconfirmed the need for additional supportive services for the students who were misclassified into the Participation Group? Was their performance on the achievement test merely a reflection of poor test taking skills? Or are these students actually in

need of services, and what factors are contributing to the nonidentification of an at-risk status? The answers to these questions may be gained by correlational data provided by teachers.

Conclusion

In summary, the findings suggest that the data from a group achievement test can be an efficient identification procedure. However, early identification has no value without a plan and without resources for treatment. Moreover, the possible adverse effects from labeling students as at-risk for academic failure should not be underestimated. The results of this investigation support the use of group achievement test data for the initial identification of students who are academically at-risk; however, further research is needed to confirm the predictive value of a group achievement test as an efficient method of identifying students in need of additional support services beyond the classroom.

References

Airasian, P. W., & Madaus, G. F. (1983). Linking testing and instruction: Policy issues. Journal of Educational Measurement, 20, 103-118.

Anastasi, A. (1982). Psychological testing (5th ed.). New York: Macmillan.

Bay, M., & Bryan, T. (1992). Differentiating children who are at risk for referral from others on critical classroom factors. Remedial and Special Education, 13, 27-33.

Bordens, K. S., & Abbott, B. B. (1988). Research design and methods: A process approach. Mountain View, CA: Mayfield Publishing.

Bruning, J. L., & Kintz, B. L. (1968). Computational Handbook of Statistics. Glenview, IL: Scott Foresman & Company.

Committee to Develop Standards for Educational and Psychological Testing. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Council of Chief State School Officers. (1987). Characteristics of at-risk students. Washington, DC: Author.

Finley, C. J. (1995). Review of the Metropolitan Achievement Tests, Seventh Edition. In J. C. Conoley & J. J. Kramer(Eds.), The twelfth mental measurements yearbook (pp. 603-606). Lincoln, NE: Buros Institute of Mental Measurements of the University of Nebraska Press.

Gresham, F. M., Reschly, D. J., & Carey, M. P. (1987). Teachers as "tests": Classification accuracy and concurrent validation in the identification of learning disabled children. School Psychology Review, 16, 543-553.

Haertel, E., & Calfee, R. (1983). School achievement: Thinking about what to test. Journal of Educational Measurement, 20, 119-131.

Hansen, J. B. (1993). Is educational reform through mandated accountability an oxymoron? Measurement and Evaluation in Counseling and Development, 26, 11-21.

Lehr, J. B., & Harris, H. W. (1988). At-risk, low-achieving students in the classroom. Washington, DC: National Education Association.

Lloyd, D. N. (1978). Prediction of school failure from third grade data. Educational and Psychological Measurement, 38, 1193-1200.

Lyman, H. B. (1991). Test scores and what they mean (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Martin, R. (1994). Section 504: Expanding schools' duties to handicapped students [Video recording]. Urbana, IL: Baxley Media Group.

Marascuilo, L. A., & Levin, J. R. (1983). Multivariate statistics in the social sciences. Belmont, CA: Wadsworth.

Mehrens, W. A., & Lehmann, I. J. (1987). Using standardized tests in education (4th ed.). White Plains, NY: Longman.

Mitchell, R. (1992). Testing for learning: How new approaches to evaluation can improve American schools. New York: Free Press.

Mitman, A. L. (1985). Teachers' differential behavior toward higher and lower achieving students and its relation to selected teacher characteristics. Journal of Educational Psychology, 77, 149-161.

National Assessment of Educational Progress. (1989). Crossroads in American education: A summary of findings. Princeton, NJ: Educational Testing Service (Report No: 17-OV-01).

Nebraska Department of Education. (1992). Title 92, Nebraska administrative code, chapter 51. Lincoln, NE: Author.

Nitko, A. J. (1989). Review of the Metropolitan Achievement Test (6th ed). In J. C. Conoley & J. C. Impara (Eds.), The tenth mental measurements yearbook (pp. 31-35). Lincoln, NE: Buros Institute of Mental Measurements of the University of Nebraska Press.

Nurcombe, B., & Partlett, D. F. (1994). Child mental health and the law. New York: Free Press.

Ogden, E. H., & Germinario, V. (1988). The at-risk student. Lancaster, PA: Technomic Publishing.

Payne, B. D., & Payne, D. A. (1991). The ability of teachers to identify academically at-risk elementary students. Journal of Research in Childhood Education, 5, 116-126.

Pedulla, J. J., Airasian, P. W., & Madaus, G. F. (1980). Do teacher ratings and standardized test results of students yield the same information? American Educational Research Journal, 17, 303-307.

Prasse, D. P. (1990). Best practices in legal and ethical considerations. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology (2nd ed., pp. 469-489). Washington, DC: The National Association of School Psychologists.

Psychological Corporation. (1994). Metropolitan achievement tests--Seventh edition: Technical manual. San Antonio, TX: Harcourt Brace & Company.

Roosa, L. W., & Biller, J. H. (1995). The intervention assistance team: A facilitator's program manual (2nd ed.). Fort Lauderdale, FL: Florida Department of Education.

Rumberger, R. W. (1987). High school dropouts: A review of issues and evidence. Review of Educational Research, 57, 101-121.

Salvia, J., & Ysseldyke, J. E. (1995). Assessment in special and remedial education (4th ed.). Boston, PA: Houghton Mifflin.

SPSS Inc. (1990). SPSS Reference Guide. Chicago, IL: Author.

Stone, B., Cundick, B. P., & Swanson, D. (1988). Special education screening system: Group achievement test. Exceptional Children, 55, 71-75.

Tabachnick, B. G., & Fidell, L. S. (1989). Using multivariate statistics: Second Edition. New York, NY: Harper-Collins.

U. S. National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform: A report to the nation and the secretary of education, U. S. department of education (SUDOCS No. ED1.2:N21). Washington, DC: U.S. Government Printing Office.

U. S. Department of Education. (1993). The chapter 1 program (SUDOCS No. ED1.402:C36). Washington, DC: U.S. Government Printing Office.

U. S. Department of Education. (1995). Individuals with Disabilities Education Act: Amendment of 1995 (SUDOCS No. ED1.202:D63). Washington, DC: U.S. Government Printing Office.

U. S. Department of Education. (1990). National goals for education (SUDOCS No. ED1.2:G53). Washington, DC: U.S. Government Printing Office.

Witt, J. C., Elliott, S. N., Kramer, J. J., & Gresham, F. M.(1994). Assessment of children: Fundamental methods and practices. Madison, WI: Brown & Benchmark.