1997

An Investigation of the Utility of Using Local Curriculum-Based Measurement Norms Versus Group Standardized Norms to Predict Students at Riskfor Academic Failure in Reading and Mathematics

Lara E. Druckman
Loyola University Chicago

LOYOLA UNIVERSITY OF CHICAGO

AN INVESTIGATION OF THE UTILITY OF USING LOCAL CURRICULUM-BASED MEASUREMENT NORMS VERSUS GROUP STANDARDIZED NORMS TO PREDICT STUDENTS AT RISK FOR ACADEMIC FAILURE IN READING AND MATHEMATICS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CURRICULUM, INSTRUCTION, AND

EDUCATIONAL PSYCHOLOGY

BY

LARA E. DRUCKMAN

CHICAGO, ILLINOIS

JANUARY 1997

ACKNOWLEDGMENTS

As I reflect on my journey of completing my dissertation, there have been many people along the way that have been exceptional and I could not have done it without them.

I wish to thank my committee for all their time and effort. In particular I wish to acknowledge:

Dr. Ronald Morgan for being my advisor, mentor, and friend.

Dr. Jack Kavanagh for his expertise in the field of statistics.

Dr. Joy Rogers for her willingness to participate on my committee.

Dr. Pamela Stecker for her knowledge of curriculum-based measurement and continued inspiration.

In addition, I want to thank my parents, family, and friends for their support and encouragement. I'd especially like to thank my mother for doing an outstanding job proofreading the final copy.

And last, but definitely not least, my fiancé, Joe, for his understanding, support, and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER ONE

INTRODUCTION

Traditional assessment practices, typically norm-referenced instruments, have been criticized for failing to generate positive impact on instruction, learning, and school practices (Jamentz, 1994). In addition, many educators question the utility, reliability, validity, and general effectiveness of many of these norm-referenced instruments. Norm-referenced tests generally are not designed to directly describe growth, but are designed to portray a student's relative standing within the population of students of the same age (Deno, 1992). Likewise, the results of many norm-referenced tests make it difficult to link them to intervention plans and evaluations.

During the 1980s and 1990s, many educators and psychologists have been interested in alternative approaches to assessing the outcomes of schooling. These new assessment methods have presented challenges to the curriculum, teaching practices, and presentation of student achievement information to policymakers and to the public (Baker, O'Neil, & Linn, 1993). Today, the focus of many schools is to move away from traditional approaches of assessing outcomes and to move toward performance-based assessment approaches.

The primary purpose of performance-based assessment is to provide information about how a student is performing relative to the curriculum in which he or she is taught.

In other words, performance-based assessment requires students to display the skills they have learned.

Over the past two decades, an increasingly popular performance-based assessment approach has surfaced. Curriculum-based measurement (CBM) is defined as a simple set of standardized procedures that teachers can use to obtain reliable and valid measures of student achievement (Deno, 1987). Mastery of the curriculum content is considered to be the basis for evaluation and remediation. Students are assessed frequently, quickly, and in a direct manner. Materials are developed directly from the local school curriculum in the basic skill areas of reading, spelling, mathematics computation, and written expression.

When CBM procedures are applied to reading instruction, a student is required to read aloud three, 1-minute probes. The number of words correctly read are recorded systematically. An analysis of a research data set collected over a 12-year period of time provides support for the notion that the number of words correctly read is an accurate measure of a student's general reading skills, including reading comprehension skills. In addition, Marston (1989) reported that CBM reading measures correlated highly with basal reading mastery tests and also with nationally standardized reading tests.

When CBM procedures are applied to mathematics instruction, a student is required to work on a sheet of mixed mathematics problems for two minutes. The number of correct digits are recorded. The number of correct digits have been reported to be moderately correlated with district criterion-referenced tests and also with nationally standardized mathematics tests (Skiba, Magnusson, Marston, & Ericksion, 1986).

The data sets collected from the reading and mathematics probes can be used to create local norms. These norms can be developed at different levels of complexity (classroom norms, school norms, and school district norms). The development of local norms provides a consistent and continuous data base that links the data collected for screening and eligibility purposes to student progress decisions (Shinn, 1988). That is to say that the norming of CBM appears to be reliable and valid. These measures can be used for screenings, eligibility determinations, instructional planning, and/or monitoring student progress. Cutoff scores can be used to determine which students require instructional modifications and/or additional academic support in order to benefit from their education.

The study to be described in the pages that follow was designed to determine the accuracy of predicting students' achievement levels using two different measures: locally normed curriculum-based measurement (CBM) procedures and group standardized testing norms. Additionally, an effort is made to document the impact of alternative assessment methods on teachers and students. This information is considered to be important and timely. It may provide support for future alternative assessment practices in education.

Participants included 16 second-, 14 third-, 15 fourth-, and 14 fifth-grade general education students selected from a suburban school district near Chicago, Illinois. During the 1993-1994 school year, each student participated in CBM Mathematics (CBM-M) and CBM Reading (CBM-R) assessment during the fall, winter, and spring. In addition, each subject was administered the Cognitive Abilities Test (CogAT), Reading Criterion

Referenced Test (CRT-R), and Math Criterion Referenced Test (CRT-M). Four teachers involved in the study completed an open-ended questionnaire designed to assess their views related to the acceptability and utility of the CBM and the CRT measures. Local norms for this study consisted of school norms that were created during the 1993-1994 academic school year. The creation of school norms followed the guidelines developed by Shinn (1989).

The following research questions were addressed:

1. Are there significant differences between the CBM locally established norms and the group norms from standardized tests for the second grade students?

2. Are there significant differences between the CBM locally established norms and the group norms from standardized tests for the third grade students?

3. Are there significant differences between the CBM locally established norms and the group norms from standardized tests for the fourth grade students?

4. Are there significant differences between the CBM locally established norms and the group norms from standardized tests for the fifth grade students?

In addition to addressing these formally stated research questions, an effort was made to determine if there were differences in the number of students identified as being at risk for academic failure when using two different measures (CBM and CRT).

1. Are there differences in the number of second grade students identified as being at risk for academic failure when the CBM outcome measures are compared to the CRT outcome measures?

2. Are there differences in the number of third grade students identified as being at risk for academic failure when the CBM outcome measures are compared to the CRT outcome measures?

3. Are there differences in the number of fourth grade students identified as being at risk for academic failure when the CBM outcome measures are compared to the CRT outcome measures.?

4. Are there differences in the number of fifth grade students identified as being at risk for academic failure when the CBM outcome measures are compared to the CRT outcome measures?

CHAPTER TWO

REVIEW OF THE LITERATURE

Assessment of students' achievement levels is the most popular and arbitrary activity that teachers perform (Gathercoal, 1995). In addition, assessment is considered to be an important tool that teachers can use to assist with the learning process. The responsibilities of the classroom teacher have increased to include new forms of assessment, school reform initiatives, the growing number of mandated assessments, increasing calls for assessment of noncognitive outcomes, and a host of entities demanding accountability.

Teacher approaches to assessment vary widely in quantity and quality. Salmon-Cox (1981) found that teachers view many achievement tests as being relatively unimportant in day-to-day decision making. Indeed, when teachers were asked whether they would miss standardized tests if these tests were abolished, only people outside of the classroom (e.g., parents, principals, school board members) reported that they would miss standardized test information. Thus, it appears that teachers do not value the information obtained through standardized achievement tests.

Historical Context

Historically, informal methods of assessment, such as direct observation, student response to teacher questioning, and scores on daily assignments dominated teacher assessment practices. Salmon-Cox (1981) found that the most common method teachers had to monitor their students' progress was through observation. Furthermore, Fuchs, Fuchs, and Warren (1982) reported that the discrepancy between actual student performance and teacher judgments of student performance based mostly on observation, proved to be statistically significant. However, the reliability and validity of teachers' informal observations of student academic performance remains unknown.

Other assessment procedures that were developed in the 1970s and 1980s which indexed student mastery of a series of objectives were mastery learning (see Block & Burns, 1976) and precision teaching (see White & Haring, 1980). Goals of both methods are to increase both instructional individualization and student achievement. Mastery learning involves programming for success or mastery, constant teacher feedback, and corrections on a prescriptive basis. Good and Brophy (1984) reported that mastery learning is successful in increasing the number of students who master basic skills. Precision teaching involves pinpointing specific behavior, recording and charting of the behavior, and changing instruction programs in response to outcome data. This method is designed to be sensitive to performance changes that can be used to evaluate program effectiveness.

Some problems, however, continue to be found with the short-term mastery focus of many of these methods. Focusing on short-term mastery monitoring makes summarizing and evaluating student progress across relatively long time periods difficult. Also, the relation between mastering many objectives and improvement on more integrated, global achievement tests remains uncertain. A considerable amount of teacher time is used to design and create tests. Unfortunately, the accuracy of many of these teacher-made tests is unknown. Finally, because different tests are developed for different students, it is difficult to compare students progress across students within the same classroom.

Anania (1982, 1983) and Burke (1984) conducted a series of studies in which student learning was examined under three different conditions of instruction: conventional; mastery learning; and tutoring. Results indicated that the average student in the tutoring groups performed about two standard deviations above the average of the control classes. Thus, the average tutored student was above 98 percent (a 2 sigma effect size) of the students in the control classes. In addition, results indicated that the average student in the mastery learning groups performed about one standard deviation above the average control classes. In other words, the average mastery learning student was above 84 percent (a 1 sigma effect size) of the students in the control classes.

This discrepancy between the differences in the final achievement measures under the three conditions is known as the "2 sigma problem". Can researchers and teachers devise groups based on teaching learning conditions that will enable the majority of students under group instruction to attain levels of achievement that can at present be reached only under highly individualized tutoring conditions? Indeed, research has been conducted regarding the possibility of combining mastery learning with two or three alterable variables in order to exceed the high level of learning that results from one-to-one tutoring. However, no variable combination has been found that has exceeded the mastery learning 1 sigma effect size or the tutoring 2 sigma effect size.

In order to help students have the potential to reach a high level of learning, researchers need to focus on mere practical and realistic conditions than the one-to-one tutoring, which is too costly and time consuming. That is to say that a practical method linking assessment to instruction needs to be established that the average teacher could learn in a brief period of time and use within the context of conventional instruction.

Another assessment method, dynamic assessment (see Vygotsky, 1962), has recently been acknowledged as an important assessment technique (Jitendra & Kameenui, 1993). Five models of dynamic assessment were designed to link assessment and instruction. Although all five models differ with regard to theoretical orientations, purposes of assessment, tasks used in the assessment, domain-specific skill evaluations, types of instruction employed, the overall goal of all five methods remains the same. The goal is "to determine and modify the reasons responsible for failure" (Jensen & Feuerstein, 1987, p. 391). The five methods include a: (a) a test-train-test assessment procedure, (b) the Learning Potential Assessment Device (LPAD), mediational assessment, (c) testing-the-limits assessment, (d) graduated prompting assessment, and (e) a continuum of assessment model-mediated and -graduated prompting.

Campione (1989) indicated that use of dynamic assessment procedures improves the predictive and prescriptive features of traditional assessment procedures by focusing on an individual's strengths and weaknesses. In fact, dynamic assessment has been used to identify students with learning difficulties and to provide information related to the effectiveness of instruction. It is important to note that the overall goal of dynamic assessment is not unique compared to other assessment methods. However, information regarding possible reasons for failure or the learner's inability to achieve is often not provided with traditional psychometric measures.

The traditional assessment practices, consisting predominantly of published, norm-referenced tests (PNRT), decontextualize problems by comparing the referred student's academic performance to the academic performance of students in national normative samples (Shinn, 1993). This comparison may have little relevance to educational decisions to be made regarding students in the local context, mostly because the students' educational experiences may have varied greatly from that of the nation.

In addition, a concern with the standardized achievement tests is that they are not designed to describe growth directly. Instead, they are designed to portray a student's relative standing within the population of students of the same age. Many commercial tests offer grade level equivalents as a way of measuring growth. However, grade equivalent scores are typically so unreliable that test publishers caution against the use of them (Deno, 1982). Likewise the use of percentile scores are useful when knowledge of a student's achievement status within his or her classroom school or within the general population is desired, but are not useful when knowledge of individual student growth in proficiency is desired.

A primary concern of PNRT is their general lack of content validity. In other words, they fail to measure directly the skills that students are expected to display. In addition, most PNRT have inadequate response formats and provide information only related to correctness. Focus is not given to error analysis. Likewise, PNRT often have an inadequate number of items distributed across a broad age and/or grade range. This arrangement does not allow for the identification of preskill deficits and/or evaluation of student progress (Shinn & McConnell, 1994).

Another reason standardized achievement tests are problematic in nature is they rely heavily on face validity. Face validity refers to whether the items on a test appear to represent what the test is supposed to measure. A primary concern with face validity is that it cannot be empirically established. Furthermore, there is a lack of agreement over definition of important skill outcomes. Messick (1980) indicated that many times if the test has face validity, it often doesn't possess construct validity. A relevant study was conducted at University of Illinois Center for the Study of Reading by Armbruster, Stevens, and Rosenshine (1977). They compared three 3rd-grade reading curricula and two standardized reading achievement tests. They found that only a small percentage of skills emphasized in the curricula were represented on the standardized tests.

A test with content validity ensures that decisions are made on the basis of what the students are expected to learn. PNRT with high content validity often lack the information for intervention planning and evaluation. In addition, in order for a test to be useful for evaluating effectiveness of the intervention, a test must be repeated and used on a frequent basis so that effective interventions are maintained and ineffective interventions are modified (Shinn & Hubbard, 1992). Often times this is not the case with PNRT.

As stated above, PNRT have received a great deal of criticism for both their lack of authenticity and their lack of utility in helping teachers improve the quality and effectiveness of their instruction (Fuchs & Deno, 1994). The failure of PNRT has lead researchers to search for assessment approaches designed to be more responsive to individual learners' potential strengths and weaknesses (Jitendra & Kameenui, 1993).

Curriculum-Based Measurement

An alternative to traditional norm-referenced assessment methods includes curriculum-based assessment (CBA) for assessing academic skill deficits (Eckert, Shapiro & Lutz, 1995). CBA includes most informal, teacher-made tests that rely on criterion-referenced measurement. As defined by Deno (1987) CBA is any set of measurement procedures that use "direct observation and recording of a student's performance in the local curriculum as a basis for gathering information to make instructional decisions" (p. 41).

Over the past two decades, an emerging alternative way of educational decision making in the school curriculum is to use direct and frequent measurements of student performance through curriculum-based measurement (CBM) (Deno, Marston & Tindal, 1986). CBM, conceptualized by Deno (1985; 1986) and Shinn (1989) can be defined as a simple set of standardized procedures that can be used to obtain a reliable and valid measure of student achievement. CBM is the result of a great deal of interest that has been generated around developing alternative assessment methods that are relevant to the primary purposes of learning and that can be used to enhance teachers' instructional planning (Fuchs, & Deno, 1994).

Curriculum-based assessment (CBA) is different from curriculum-based measurement (CBM) in that it refers to a variety of approaches to assessment that rely on gathering information on performance in the curriculum. Fuchs & Deno (1991) indicated that CBM differs from CBA in that it does not rely on task analysis, subskill analysis, or mastery learning. In addition, CBM differs from CBA in that it provides a data base for making educational decisions beyond the initial assessment phases (Shinn & Hubbard, 1992). Although there are several curriculum-based assessment models, there are three similar features: student proficiency must be sampled in material from the school's curriculum, assessments must recur over time, and information must be used to formulate instructional decisions (Deno, 1985; Fuchs & Deno, 1991; Gickling & Havertape, 1981; Shapiro, 1990; Shinn, 1989; Fuchs & Deno, 1994).

CBM first started at the University of Minnesota Institute for Research on Learning Disabilities (IRLD) in an effort to decrease the separation between measurement and instruction (Deno, 1985). The primary purpose of this study was to develop measurement procedures that teachers could use to help decide whether to modify a student's instructional program. The researchers in this study created a set of procedures that included the following four characteristics; measures would be reliable and valid, measures would be easy to administer, measures would be designed to enable repeated and frequent administration, and measures would be time efficient and cost effective.

CBM, a performance-based assessment approach, has become increasingly popular and has been used primarily to provide special education teachers with a method of evaluating the effectiveness of their instructional interventions with individual students (Shinn & Habedank, 1992). One reason for the popularity of CBM is that the procedure allows teachers to determine directly the extent to which a student is learning what is taught. In addition, the fact that CBM is conducted within the context of the regular curriculum of the local school is appealing to many educators and school boards, because it helps preserve the sense of local control (Fuchs, & Deno, 1994).

Standardized CBM procedures have been developed for measuring growth in reading, spelling, written expression, and mathematics computation. The standardized procedure on which the teacher relies consists of sampling the curriculum to create the CBM probes, administering and scoring the probes, analyzing the students' performance, and formulating instructional decisions. CBM helps to provide a context for problems by determining if a discrepancy exists between the referred student's academic performance and the performance of typical students in the local school community. The larger the discrepancy, the more severe the problem.

In contrast to standardized achievement tests, the CBM procedures permit the comparison of the referred students' academic performance to the performance of typical peers who have had, on the whole, similar instructional opportunities, curricula, and learning experiences. Likewise, CBM allows for repetitive and direct measurement of academic skills, thus allowing for frequent assessment and growth monitoring. In addition, CBM allows the teacher to determine directly the extent to which a student is learning what is being taught. Thus, the CBM procedures helps teachers index student progress, evaluate the effectiveness of their instruction, and design better programs (Fuchs & Fuchs, 1986).

As reported by Fuchs and Fuchs (1991), CBM is now used for a variety of psychoeducational assessment purposes: to formulate student goals, to determine when instructional adjustments are necessary to increase the probability of goal attainment, to identify specific strategies to enhance instruction, and to monitor the appropriateness of student goals and adjust them as necessary. In addition, Shinn & Habedank (1992) discuss the utilization of CBM for problem identification and problem certification decisions (i.e., eligibility for special education). In order to determine eligibility for special education Jenkins, Deno, and Mirkin (1979) proposed that frequent measurement of skills in an academic area be compared with minimal acceptable performance in that area (Marston, Deno, & Tindal, 1984).

There are six basic advantages reported to be associated with the use of this standardized procedure. First, the time consuming burden of developing measurement procedures is removed from the teacher. Second, the process for measuring student performance within an academic area remains constant across time for each pupil and across different pupils. Third, the need for clear effective communication of student performance is achieved. Fourth, the procedures are sensitive to growth in student performance over relatively short periods of time. Fifth, the process is cost-effective. Sixth, the teacher can be confident in the meaningfulness and accuracy of the scores in determining students who are at risk of academic failure and who may require a change in the instructional program they receive.

Teacher Acceptability. Eckert, Shapiro, and Lutz (1995) found that when comparing teachers' ratings of both CBM and PNRT, CBM was consistently rated as a more acceptable method of assessment than PNRT. In addition teachers viewed CBM as an effective and appropriate approach in assessing academic skills problems. The teachers also indicated that they thought CBM would be effective for identifying children's problems, applicable for a variety of children and academic problems, and beneficial for students.

A similar study comparing group achievement measures, teachers' ratings, and CBM was completed by Wilson, Schendel, & Ulman (1992). This study also found that all three of the above mentioned tools appear to have utility as alternative screening or assessment measures for children in need of special and remedial services. As reported by Mirkin, Fuchs, & Deno (1982), teachers indicate that CBM is useful for pinpointing accountability, for providing feedback to and motivating students, and for formulating Individualized Education Program (IEP) goals, objectives, and monitoring procedures.

Likewise, in 1994, Shapiro and Eckert investigated the acceptability of CBM to PNRT among school psychologists. The Assessment Rating Profile (ARP; Kratochwill &

VanSomeren, 1984) was used and results indicated that CBA had significantly higher acceptability ratings than did norm-referenced assessment procedures.

Limitations. Despite the many positive aspects of CBM, some limitations have been noted. Some researchers question the norming process of CBM (Mehrens & Clarizio, 1993). They claim that CBM relies only on local norms and thus is an inadequate procedure for making national norming inferences. In addition, Mehrens & Clarizio (1993) question the reliability of CBM. These critics cite several studies that used heterogeneous groups in which the standard error of measurement and standard error of difference scores were not reported.

Validity concerns with CBM have also surfaced. The studies reviewed by Marston (1989) indicated that most CBM criterion-related validity studies are based largely on PNRT. Given the argument of the CBM proponents that tests are more useful if tied to the local set of objectives, it is difficult to know why studies almost invariably use PNRT as a criterion in their criterion-related studies (Mehrens & Clarizio, 1993). Researchers have also noted that construct/decision validity remains questionable at this time due to scant evidence related to diagnostic decision making.

Another area of concern that has received considerable criticism is related to the treatment utility of CBM. Many conclude that an advantage for CBM is knowing when to modify instructional planning. However, CBM fails to be prescriptive with respect to what to change and how best to instruct the student. This overall failure to demonstrate treatment utility is also evidenced with nationally standardized tests.

Local Norms. The idea of local norms has been around for a long time. However, recently local norms as a decision-making standard for educators has received a great deal of attention (Shinn, 1989). Anastasi (1988) stated "local norms are more appropriate than broad national norms for many testing purposes such as . . . comparison of a child's relative achievement in different subjects or the measurement of an individual's progress over time" (p. 98). In addition, local norms appear to decrease bias (Oakland & Matuxzek, 1977) and offer more information, especially in cases concerning minority students (Elliott & Bretzing, 1980).

Deno (1985, 1986) suggests that CBM is based upon two major premises: (a) assessment and decision making are curriculum referenced, and (b) special education decision-making is both individually and normatively referenced. Thus, the pupil's academic progress is indexed against local normative performance in the curriculum. In other words, local norms provide an index of the expectations of the regular education environment.

Developing local norms requires developing a representative of grade-level curricula materials for grade levels to be tested, establishing a normative sampling plan, training the collectors, collecting the data, and summarizing the data. The development of local norms is feasible because the CBM data is cost-efficient and time-efficient.

As indicated by Shinn (1988), local norms provide a consistent and continuous data base that links the data collected for screening and eligibility purposes to student progress decisions. CBM procedures and local norms are often used by school districts to make special education screening decisions. For example, if a referred student is sufficiently different from general education peers, further assessment is warranted.

An advantage of creating local norms is the meaning they provide to any particular score that changes as a function of grade, material, and time of testing (Shinn, 1988). In addition, local norms can be used for many special education decisions including, screening, eligibility, writing IEP objectives, monitoring progress, periodic and annual reviews, and program evaluations.

Despite the advantages of using locally generated norms, there are some disadvantages for using local norms that need to be considered. The greatest disadvantage is that the screening and eligibility procedures have the potential to be used for labeling children as disabled without connection to developing more effective programs. Another potential concern with local norms is the acceptance of mediocrity when local normative performance is accepted as a goal for all children. Yet another potential disadvantage is the possibility of being perceived as advocating the general education curriculum as "the curriculum" (Shinn, 1988).

Screening & Eligibility. It is estimated that schools refer an average of 5% of the general education population for special education on a yearly basis (Shinn, 1989). The potential of making a biased decision is increased if there is not a systematic process for making that decision and also if there is little control over the purpose of the teacher referral. A study by Shinn, Tindal, and Spira (1987) revealed that all referrals need to be evaluated in a timely and systematic manner to try to eliminate the part teacher tolerances and biases play in the referral process.

In the screening and eligibility process using CBM, referred students are compared first to the normative performance of grade-level peers on grade-level curricular tasks. Shinn (1989) suggests two different methods to determine which students require further evaluation for determining academic difficulties: the discrepancy ratio and the percentile rank. In terms of special education eligibility, Shinn recommends when using CBM with the discrepancy ratio method, a cutting score of 2.0 be used. To use the discrepancy ratio method to determine if a significant discrepancy exists, the peer median is divided by the referred student's median score. If that discrepancy is greater than 2.0 further assessment may be warranted. When using CBM with the percentile method, Shinn (1989) recommends a 10th percentile cutting score be used. To use the percentile rank method the percentile score which corresponds to the referred student's median score is found. If the referred student's median score falls at or below the 10th percentile the student's performance is considered significantly discrepant and may warrant further assessment.

A number of CBM methods have been used to determine students' eligibility for services for mild disabilities. Within PL 101-476, eligibility criteria are based upon dimensions of variability from achievement expectations for average students. Indeed, these criteria can be modified as a result of social, economic, and political factors.

As reported by Marston & Magnusson (1985), by using CBM, the results of a district-wide screening process conducted with all students referred for special education reduced the number of eligibility assessments by approximately 40%. In addition, Marston, Mirkin, and Deno (1984), contrasted a weekly CBM screening in spelling, reading, and written language with a traditional teacher referral procedure. Results indicated that the number of special education referrals from the two different procedures were similar. Indeed, they found very few differences in the types of pupils that were identified to be served.

In the Minneapolis Public Schools, students can be eligible for special education if students perform in the range of average students two years below their current placement on the corresponding grade-level curriculum materials (Marston, & Magnusson, 1985). Likewise, students in the Pine County Special Education Cooperative are considered for eligibility if their performance is at half the rate of peers on grade-level materials, given the other exclusionary components of 101-476 (Germann & Tindall, 1985).

A study by Shinn, Ysseldyke, Deno, and Tindal (1986) concluded that CBM measures might be of value in the identification of students in need of services. In addition there has been extensive use of CBM data in screening students who are at risk for school failure (Allen, 1989; Shinn, Tindal, & Stein, 1988).

Technical Adequacy. Regarding the technical adequacy of CBM, many researchers have concluded that standardized CBM procedures are valid and reliable (Marston & Magnusson, 1988; Ysseldyke, Pianta, Christenson, Wang, & Algozzine 1983; Deno, 1985; Fuchs & Fuchs, 1986; Shinn, 1989). In general, the reliabilities from five different studies were found to be sufficiently high (Marston, 1989). However, it should be noted that two of the five studies used students across grades, thereby creating very heterogeneous groups and possibly inflating the reliability estimates. In addition, it should be noted that standard errors of measurement and standard error of difference scores were not reported in these studies (Mehrens & Clarizio, 1993).

In sum, even though the available evidence lends support to the adequacy of CBM reliability, the adequacies of CBM reliabilities should be accompanied by cautionary and explanatory qualifiers. This is due to the extreme heterogeneity of the group in some of the studies, the failure to report standard errors of measurement as well as standard error of difference scores, and also the inattention to the effects of heterogeneity on reliability estimates.

The material involved in CBM is derived from the actual curriculum of the local school, and thus it is assumed that one desires a test with high local curricular validity. The concern is that CBM is primarily concerned with basic skills assessment (i.e., reading, writing, spelling, and mathematics computation problems). The actual behavior sampled from CBM is even more limited than that (Mehrens & Clarizio, 1993). For example, in the area of reading, the number of words read correctly is examined, in mathematics, the amount of correct digits is examined, in written expression, the number of words written correctly is examined, and in spelling, the number of correct letter sequences is examined. Mehrens and Clarizio (1993) indicate that the local curriculum is far broader than the domains sampled by CBM. Indeed, the measures sample the curriculum, but the sample is obviously not representative.

Deno (1985) reported that criterion validity studies revealed all of the reading curriculum-based measures were highly correlated with performance on PNRT. However, eight of the fourteen studies he reviewed grouped students across several grade levels. Jenkins and Jewell (1993) found that indeed criterion validity coefficients were generally smaller within grades than across grades. In addition they found that the coefficients decreased as the grade levels increased. They concluded that the concurrent validity may depend on the student's grade level.

Studies examining the criterion-related validity for CBM reading passage are summarized by Marston (1989). These studies correlated the CBM measures with Reading PNRT. The results of these studies indicated that Reading CBM was found to highly correlate with Reading PNRT. The three group achievement tests reported as criterion variables were the Stanford Achievement Test (SAT; Madden, Gardner, Rudman, Karlsen, & Merwin, 1973), Science Research Associates reading subtests (SRA; Naslund, Thorpe, & Lefever, 1978), and the California Achievement Test (CAT; CTB/McGraw-Hill, 1985). In the studies mentioned above, the sources of reading curricula for CBM measures included the following basal reading series: Allyn-Bacon (Rudell, Monson & Reid, 1978); Ginn 720 (Cymer, Green, Gates & McCullough, 1976); Ginn Reading Series (Clymer & Fenn, 1979); and Houghton-Mifflin (Durr, Lepere, Dean, Glaser & Lewis, 1976).

Also, Deno, Mirkin, & Chiang (1982) examined criterion validity coefficients for different criteria of Reading CBM (i.e., cloze, word meaning, isolated oral word reading, and passage oral reading measures). They found that listening to students read aloud from their basal reader for 1 minute was a valid measure of their reading skill. In addition the correlation coefficients ranged from .73 to .91, with most coefficients above .80. In addition, Fuchs, Fuchs & Maxwell (1988) compared criterion-related validity coefficients for additional criteria for CBM (i.e., cloze, retell, question answering, and passage oral reading measures). This study also concluded that reading aloud from a text demonstrates the strongest relation to widely used criterion measures of reading. Thus, these findings indicate why oral reading fluency is used more frequently with Reading CBM.

In addition to the many studies comparing the relation between CBM reading fluency and reading skill, a third factor was added to the equation. When adding the teachers' holistic rating of the students' reading ability, Fuchs and Deno (1981) found that indeed reading fluency measures were highly related to teachers' judgment of student reading proficiency. Moreover, a study completed by Marston and Deno (1982) indicated that there was a stronger relationship between oral reading fluency and teacher holistic ratings of reading skills versus teacher ratings with published achievement tests and their actual reading placement in the curriculum. Such findings also demonstrate reading fluency's criterion-related validity.

In addition to criterion-related validity, other methods can be used to judge the validity of a measure. Deno, Marston, Shinn, & Tindal (1983) have provided evidence supporting discriminant validity. Marston, Deno, Mirkin, Lowry, Sindelar, & Jenkins (1981) have provided evidence supporting longitudinal change. Deno (1985, 1986) has provided evidence supporting sensitivity to reading programs, and Fuchs & Fuchs (1986) have provided evidence supporting treatment validity.

Studies examining the criterion-related validity for CBM mathematics probes are summarized by Marston (1989). These studies reported by Marston correlated the CBM measures with Mathematics PNRT (Skiba, Magnusson, Marston, & Erikson, 1986). The three group achievement tests reported as criterion variables were the MAT Operations and the MAT Problem Solving Test (Durost, Bixler, Wrightsone, Prescott, & Balow, 1971) and the District CRT Basic Mathematics Concepts tests. In the studies discussed above, the sources of mathematics curricula included probes composed of addition, subtraction, multiplication and division problems specific to grade-level curricula.

These validity studies with mathematics are not as promising as the studies with reading. Studies completed by Skiba, et. al. (1986) indicate that few correlations exceed .60 and it appears as if the validity coefficients increase as the age of the subjects increases. Skiba et. al. (1986) offers two reasons why these lower than expected findings were found. The first reason being that there is a concern for using the published mathematics tests as a criterion measure because many mathematics tests have limited content validity. In addition, the researchers found that when reading skills were added to the prediction equation, the coefficients obtained improved significantly. Thus, possibly indicating, that the criterion mathematics test could also be measuring more than just mathematics computation.

Recapitulation

It is probably fair to say that most teachers require a simple, valid, and efficient procedure to monitor student progress in order to make judgments regarding the effectiveness of their efforts to individualize their instruction. In addition, teachers strive to determine whether students are learning what is being taught.

Informal teacher observations of student performance require no additional time and/or materials from the teacher while they are teaching. However, the reliability and validity of these informal nonstandardized teacher assessments remains questionable and controversial.

Many of the mastery tests that have been developed by publishers to help standardize teacher judgments about student progress have been reported to be technically inadequate. Moreover, these mastery tests are often given infrequently, and make it difficult for continuous monitoring and evaluation of student growth (Deno, 1985). Likewise, assessment methods such as, mastery learning and precision teaching, which index student mastery of a series of objectives, continue to demonstrate problems with their short-term mastery focus. The focus on short-term mastery monitoring makes summarizing and evaluating student progress difficult.

Results from PNRT often are difficult to link to interventions and seldom enhance teachers' ability to monitor academic progress over time (Shapiro, 1989; Shinn, Nolet, & Knutson, 1990). Additionally, the time required to administer these tests varies among tests, but, for the most part, administering PNRT takes considerably more time than giving alternative assessments. It should be noted that with commercially distributed achievement tests, even if they are found to be technically adequate, the school is paying for a complex and time-consuming procedure that yields a norm-referenced score. The test gives no information related to the student's competence in the local school curriculum, nor will the test give information regarding how the student is performing in the curriculum relative to the student's classmates.

Despite a number of unfavorable issues associated with the use of CBM procedures (Baker, O'Neil, & Linn, 1993), CBM has become a rather popular alternative procedure focused on assessing academic skill problems. CBM procedures appear to be cost-effective because no additional materials need to be purchased. Fuchs, Wesson, Tindal, Mirkin, & Deno (1982) conducted research regarding the allocation of teacher and student time to determine the amount of time required for CBM. Their results indicated that in order to maximize the efficiency of CBM, teachers must be carefully trained and prepared.

There is also some evidence that the time taken to frequently test student performance in the curriculum can actually lead to improved student achievement (Mirkin, Deno, Tindal, & Kuehnle, 1982). In fact, researchers have demonstrated that implementation of CBM procedures results in greater academic gains in reading (Fuchs, Fuchs, Hamlett, & Ferguson, 1992) and mathematics (Fuchs, Fuchs, Hamlett, & Stecker, 1990) only if teachers use information to make instructional changes.

In addition, CBM has proven to be a direct and systematic approach for assessing and monitoring academic achievement. CBM procedures can be utilized to match student performance with instructional requirements and facilitate progress monitoring of academic skills (Fuchs, Fuchs, Hamlett, & Stecker, 1990). Indeed, Thurlow and Ysseldyke (1982) surveyed teachers and found that they preferred assessment methods measuring specific academic skills. Likewise, Eckert et. al. (1995) found that teachers viewed CBM as an effective and appropriate approach in assessing academic skill problems.

CHAPTER THREE

METHOD

As noted above, this investigation was designed to determine the accuracy of predicting a student's achievement levels using two different measures (locally normed curriculum-based measurement procedures and group standardized testing norms). The study was designed to focus on the following two goals: (a) to determine the relationship between locally generated CBM norms and standardized group norms, and (b) to determine the feasibility of predicting students at risk for academic failure through using locally generated norms.

Hypotheses

The following null hypotheses were tested:

1. There are no significant differences between the CBM locally established norms and the group norms from standardized tests for the second-grade students.

2. There are no significant differences between the CBM locally established norms and the group norms from standardized tests for the third-grade students.

3. There are no significant differences between the CBM locally established norms and the group norms from standardized tests for the fourth-grade students.

4. There are no significant differences between the CBM locally established norms and the group norms from standardized tests for the fifth-grade students.

In addition to testing the four null hypotheses listed above, an effort was made to address four additional research questions:

1. Are there differences in the number of second-grade students identified as being at risk for academic failure when the CBM outcome measures are compared to the CRT outcome measures?

2. Are there differences in the number of third-grade students identified as being at risk for academic failure when the CBM outcome measures are compared to the CRT outcome measures?

3. Are there differences in the number of fourth-grade students identified as being at risk for academic failure when the CBM outcome measures are compared to the CRT outcome measures?

4. Are there differences in the number of fifth-grade students identified as being at risk for academic failure when the CBM outcome measures are compared to the CRT outcome measures?

Setting

During the 1993-94 school year the enrollment of Community Consolidated School District #59 was 6,156 students. District #59 is comprised of 13 elementary buildings and 3 junior high buildings. The district provides instruction for preschool students and students from Grades K-8. White non-Hispanics, Black non-Hispanics, Hispanics, Asian/Pacific Islanders and Native Americans (American Indians/Alaskan Natives) are the major racial-ethnic groups attending Illinois public schools. The enrollment at District #59 consists of 70.6% White, 3.5% Black, 13.9% Hispanic, 11.7% Asian/Pacific Islander, and 0.2% Native American. This information was obtained from the Illinois State Board of Education 1993-1994 School Year Report Card.

The district enrollment consisted of 7.3% low-income students. These students were defined as those from families that receive public aid, those living in institutions for neglected or delinquent children, those being supported in foster homes with public funds, or those eligible to receive free or reduced-price lunches. In addition, 9.6% of the students were considered to be Limited-English-Proficient, and were eligible for bilingual education.

For the 1993-1994 school year, District #59 had a 95.6% attendance rate. This is defined as the percentage of students who attend school every day. The student mobility rate was 19.0% and is based on the number of students who enroll in or leave a school during the school year. The district had 0.0% chronic truants (i.e. those who were absent from school without valid cause for 10% or more of the last 180 school days).

Sample

Students. Subjects were 16 second-, 14 third-, 15 fourth-, and 14 fifth-grade general education students selected from Admiral Byrd Elementary School (Community Consolidated School District #59) located in Arlington Heights, Illinois. Originally, 20 random subjects were chosen from each grade level. However, due to absences and some children moving away, there were fewer than 20 subjects at each grade level.

The gender distribution for each of the grade levels was 75% male and 25% female for the second-grade students, 71% male and 29% female for the third-grade students, 60% male and 40% female for the fourth-grade students, and 57% male and 43% female for the fifth-grade students. Table 3.1 presents a comparative overview of the demographic characteristics of the subjects.

Table 3.1 – Comparative Summary of Targeted Student Demographics

Characteristic	Grade 2	Grade3	Grade 4	Grade 5
Race				
White	15	13	15	11
Black	0	0	0	0
Hispanic	0	1	0	1
Other	1	0	0	2
Gender				
Male	12	10	9	8
Female	4	4	6	6
Total	16	14	15	14

Teachers. Four teachers were chosen carefully and asked to complete a questionnaire regarding their thoughts and feelings related to the acceptability and feasibility of using the reading and mathematics curriculum-based measurement measures and the reading and mathematics criterion-referenced test measures. These four subjects were chosen because they all had experiences preparing, administering, scoring, and interpreting both the CBM and the CRT results. The four subjects selected for inclusion in the study were an assistant principal/mathematician, a reading clinician, a student resource assistant, and a learning disabilities teacher.

Instrumentation

Reading Criterion-Referenced Test (CRT-R). The Reading CRT (CRT-R) was a criterion-referenced, group administered, achievement test designed for Grades 1-12. The test was used districtwide and developed by the book publishers (Harcourt, Brace, & Jovanovich - HBJ). The CRT was considered to be the End-of-Book test for each level of the HBJ Reading Program. The overall purpose of this test is to measure each student's progress through the basic reading curriculum. The test results are used to provide information on the ability to read and comprehend grade-level material. In addition, the results are used to provide general areas of strengths and/or weaknesses so teachers can plan the most appropriate reading program the following year.

The reading test contained several different subsections depending on the grade level (i.e., decoding, vocabulary, comprehension, study skills, and literature). It should be noted that only the total reading composite scores were used in the data analysis. The CRT-R is generally administered in the month of May.

Information regarding the standardization and technical adequacy of the CRT-R was provided through the Harcourt Brace & Company (HBJ) by Beck Evaluation and Testing Associates, Inc. During the 1987-1988 school year, HBJ conducted a field testing of the End-of Book (CRT) tests for each level of the HBJ reading program. Over 5000 students from Grades K-8 from 12 school systems across the country were included in the standardization study. However, only 1188 students took part in the End-of-Book test. The school systems that participated in the study were those that had adopted the HBJ reading program for districtwide use, used the Unit and/or End-of-Book tests as part of their program, and were willing to share the results with the publisher.

Test reliability was found to be within a range from .86 to .97. This finding indicates that the End-of-Book scores are highly reliable. However, these reliability indices are not surprising, because most of these tests are rather long (i.e. most include over 100 questions).

Test validity was assessed relative to teacher judgments of student reading levels. Participating teachers were asked to fill out a questionnaire in which they provided an estimate of each student's reading level, rated various portions of the tests, and indicated the use(s) they made of the CRT tests. Those students judged by their teachers as being the better readers invariably outperformed their peers who were teacher-named as being weak readers. Participating teachers also were asked to predict the scores each of their students would receive on the test. Results indicated that teacher estimates agreed quite closely with actual student scores in terms of rankings. That is, those students predicted to be the best, the weakest, and so on were actually ranked that way. Again, the investigators built a case for the notion that this is another way to support the validity of the End-of-Book tests.

Mathematics Criterion-Referenced Test (CRT-M). The Math CRT (CRT-M) was a criterion-referenced, group administered, achievement test designed for Grades 1-12. The test was used districtwide and developed through the district office. The overall purpose of this test was to measure and track each student's progress through the basic mathematics curriculum. The test results were used to provide information on the ability of the student to display appropriate mathematical skills for the grade level. In addition, the results were used to provide general areas of strengths and/or weaknesses so teachers can plan the most appropriate mathematics program the following year. The mathematics test contained several different subsections depending on the grade level (i.e., addition, subtraction, problem solving, time, money, measurement, graphing, etc.). Once again, it should be noted that only the total mathematics composite scores were used in the data analysis. The CRT-M is generally administered in May, however, on a different day than the CRT-R.

There appears to be no standardization and technical adequacy information regarding the CRT-M. This is due to the CRT-M being developed by mathematicians, whom are employees of School District #59. Thus, the test is individualized with respect to District #59. Unfortunately, there are no reliability or validity coefficients available that could be used to determine whether the test measures what it purports to measure and whether the measure is consistent over time. The test developers reportedly made an attempt to design the CRT-M to relate to the mathematics curriculum (i.e., the standard for the district).

Cognitive Abilities Test (CogAT). The CogAT is a standardized, group administered, norm-referenced test designed to measure a student's ability. This test provides an appraisal of the level and pattern of cognitive development of students in grades K-12. It reportedly measures abilities that are associated with problem solving in a variety of contexts. Two editions of CogAT Form 5 (Levels 1-2 for grades K-3 and Levels A-H for grades 3-12) were used. The CogAT yields a nonverbal, verbal, and quantitative score. It should be noted that both the verbal and nonverbal scores were used in the data analysis. The verbal battery was selected for use because it appears that the test plays an important role in predicting reading and oral comprehension abilities. In order to obtain a better estimate of mathematical ability, the nonverbal battery was used instead of the quantitative battery. This is due to the fact that the nonverbal battery requires no reading, whereas the quantitative battery does.

Regarding the standardization of the CogAT, it was administered under uniform conditions to a representative sample of students from each grade level during the spring of 1992. The score distributions that were obtained from the national standardization process are the norms that provide a basis for interpreting student performance.

Normative data collected at the time of standardization allow the different ability areas to be placed on a common score scale, so that a score in one area can be compared with scores in the other two areas. The common score scale for CogAT is called the Universal Score Scale.

The national standardization sample for CogAT consisted of approximately 160,000 students in Grades K-12 and included public, Catholic, and private non-Catholic schools. Table 5.25 (Raw Score Summary Statistics CogAT 5 - 1992 National Standardization, located in the book Riverside 2000: Technical Summary 1 contains the means, standard deviations, standard errors of measurement, and reliability coefficients (KR-20) for spring raw scores. The average reliabilities for the Verbal, Quantitative and Nonverbal Batteries for CogAT 1 and 2 (grades K-2) are .83, .89, and .912 respectively. For levels A-H (grades 3-12), the reliabilities average .94 for Verbal, .92 for Quantitative, and .95 for Nonverbal for spring.

Curriculum-Based Measurement Reading Passages (CBM-R). The reading passages subtest of the CBM instrument was developed using the procedures outlined by Shinn (1989). Three passages at each administration were randomly selected from the HBJ reading series for each grade level. (see Appendix A for sample grade-level CBM-R passages). Selected passages did not include poems, pictures, considerable dialogue, many proper nouns, unusual words, or decoding exercises. It should be noted that each student was presented individually the same three designated grade level readings.

The directions consisted of the examiner telling the student: "When I say 'start', begin reading aloud at the top of this page. Read across the page [demonstrated by pointing]. Try to read each word. If you come to a word you do not know, I'll tell it to you. Be sure to do your best reading. Are there any questions?" The examiner then said to the student, "Start." Students read each passage for 1 minute. The examiner followed along on his or her copy of the story, marking the words that were incorrectly read. If a student struggled with a word for 3 seconds, the examiner told the student the word and marked it as incorrect. The examiner then took the number of words read minus the number of words incorrectly read to obtain the number of words read correctly (WRC). The total score was the median of the scores across the three passages.

Research indicates that the development of reliable and valid measures of the basic skills proceeded in a step-by-step manner. First, an extensive review of the literature was conducted. Second, research teams met several times in order to review the potential measures with regard to the established characteristics. Third, the measures that appeared to meet most criteria were field tested for their criterion-related validity. Fourth, reliability studies were conducted. Fifth, studies of logistics of measurement were conducted (Marston, 1989).

In the area of reading, reliability estimates were found to be highly positive. Using test-retest intervals of 1 to 10 weeks, test-retest reliability coefficients ranged from .82 to .97, with most correlations being above .90. In addition, parallel form estimates ranged from .84 to .96, with most correlations being above .90. Similarly, interrater agreement coefficients were found to be .99 (Marston, 1989). The reader is referred to Table 2.2 in Curriculum-Based Measurement: Assessing Special Children (Shinn, 1989) for a nicely crafted in-depth comparative summary of the studies conducted.

Deno, Mirkin, and Chiang (1982) correlated five different measures of reading that potentially could be employed to monitor students' progress on a frequent basis with generally accepted PNRT. The criterion measures chosen for the first study were the Stanford Diagnostic Reading Test (Karlsen, Madden, & Gardner, 1975), the Woodcock Reading Mastery Test (Woodcock, 1973), and the Peabody Individual Achievement Test (Dunn & Markwardt, 1970). Results of these early studies indicated that the correlations among the various measures ranged from.73 to .91, with most of the coefficients being above .80. Latter studies which correlated oral reading fluency with different published measures of global reading skills ranged from .63 to .90, with most coefficients being above .80. In addition, other studies concluded that the curriculum-based reading measures shared more variance with those basal mastery tests that were correlated highly with general measures of reading skills than with those that were less related to other measures of reading ability (Marston, 1989).

Curriculum-Based Measurement Mathematics Probes (CBM-M). The mathematics probes subtest of the CBM instrument was also developed based upon procedures outlined by Shinn (1989). A probe consisted of approximately 20 random mathematical problems (see Appendix B for sample grade-level CBM-M probes). Second grade probes consisted of addition and subtraction calculations. Third grade probes consisted of addition, subtraction, and multiplication calculations. Both the fourth and fifth grade probes consisted of addition, subtraction, multiplication, and division calculations. The mathematics probes were ordered from the University of Oregon, Education Department. These probes were administered to the students in the classroom as a large group.

The directions consisted of the examiner telling students: "The sheets on your desk are mathematics facts. There are several types of problems on the sheet. Some are addition, some are subtraction, some are multiplication, and some are division [as appropriate]. Look at each problem carefully before you answer it. When I say 'start,' turn them over and begin answering the problems. Start on the first problem on the left on the top row [point]. Work across and then go to the next row. If you can't answer the problem make an 'X' on it and go to the next one. Are there any questions?" The examiner then said to the students, "Start." After 2 minutes, performance was scored in terms of number of correct digits. (e.g., If a student's answer was 2765 to a problem requiring an answer of 2865, he or she was awarded 3 of 4 correct digits.) The total score consisted of the sum of correct digits across problems.

In the area of mathematics, a number of reliability studies have been conducted (Tindal, Germann, Deno, 1983; Tindal, Marston, & Deno, 1983). Both test-retest and parallel form estimates were reported to be high, ranging from .93 to .98. In addition, the interscorer agreement was high, ranging from .93 to .98.

The limited mathematical technical adequacy data is provided by Skiba et. al. (1986). Overall, very few correlations exceed .60 and the median correlation is .425 with Metropolitan Achievement Test Problem-Solving (MAT) and .54 with MAT Math Operations. Two hypotheses were proposed by Skiba et al. (1986) in an effort to explain the lower validity correlations. First, there is a concern regarding the suitability of published mathematics tests as a criterion measure due to the limited content validity of many mathematics tests (Freeman, Kuhs, Porter, Floden, Schmidt, & Schwille, 1983). Second, Skiba et al. (1986) found in his studies that the coefficients were significantly improved when each student's reading skills were included in a prediction equation. Given this finding, it was concluded that the criterion mathematics tests could be measuring more than just mathematical computation skills.

Curriculum-Based Measurement Norms. Locally developed curriculum-based norms can be developed at three different levels (classroom norms, building norms, and school district norms). For this study, building norms were developed for each grade level at Byrd School. Table 3.2 displays the measurement net that was used to create the reading portion of the building norms. The measurement net identifies the grade-level materials that were administered for each grade level.

Table 3.2

Measurement Net for the Reading CBM Measures

Grade	Reading	Administration Time
2	Harcourt, Brace, & Jovanovich Basal Reader; Weathervanes	3 passages, each read for 1 minute
3	Harcourt, Brace, & Jovanovich Basal Reader; Celebrations	3 passages, each read for 1 minute
4	Harcourt, Brace, & Jovanovich Basal Reader; Crossroads	3 passages, each read for 1 minute
5	Harcourt, Brace, & Jovanovich Basal Reader; Skylines	3 passages, each read for 1 minute

In addition, Table 3.3 displays the measurement net that was used to create the mathematics portion of the building norms. The measurement net outline is used to identify the grade-level materials that were administered for each grade level.

Table 3.3

Measurement Net for the Mathematics CBM Measures

Grade	Mathematics	Administration Time
2	Mixed Probe (+, -)	2 minutes
3	Mixed Probe (+, -, X)	2 minutes
4	Mixed Probe (+, -, X, ÷)	2 minutes
5	Mixed Probe (+, -, X, ÷)	2 minutes

Once all the data were collected, scores were organized by grade level using means, medians, and standard deviations. Box plots were created similar to the ones seen in Figures 3.1 and 3.2. This graphic format represents the display of the range of average scores (i.e., from the 25th to the 75th percentile) across grades. Thus, the boxes in the figures represent the range of scores of typical students in the general education classroom in grade-level curricular materials for the spring norming period. The dark horizontal line represents the median performance for each grade level in the academic area specified.

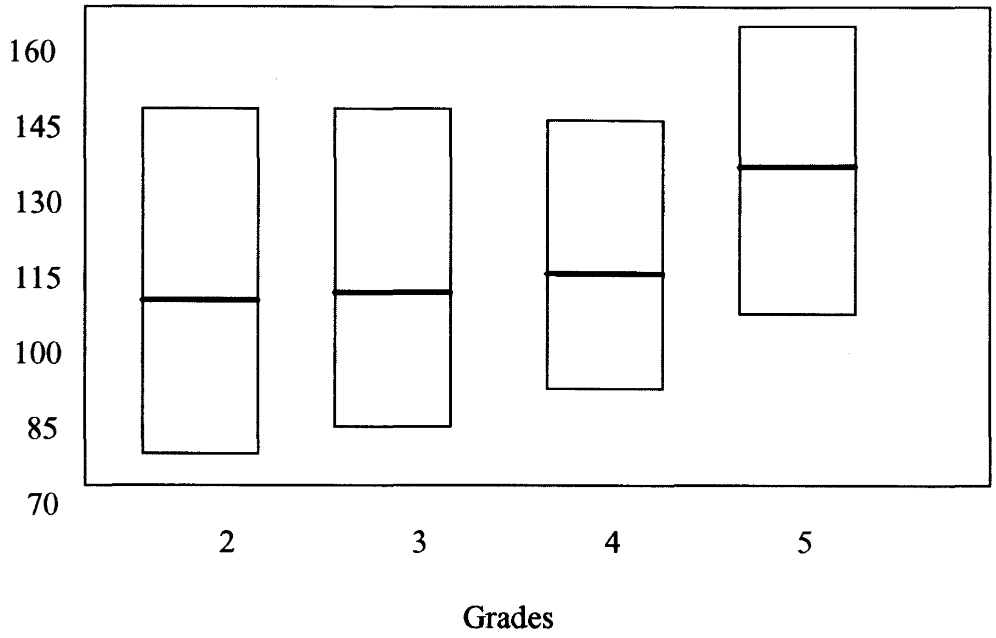

FIGURE 3.1. The range of reading scores during the spring norming period for grades 2 through 5.

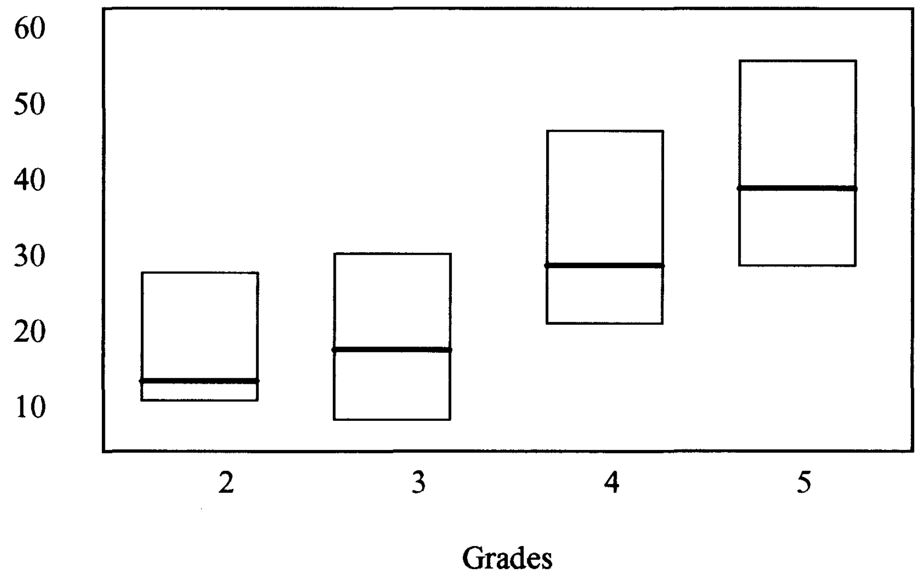

FIGURE 3.2. The range of mathematics scores during the spring norming period for grades 2 through 5.

Survey. The purpose of the survey was to examine the teachers' thoughts and feelings on the acceptability and feasibility of two assessment methods (CBM and CRT) (see Appendix C). The investigator developed a survey in which the teachers were asked to identify both advantages and disadvantages of the acceptability of the two assessment procedures. The survey consisted of 10 open-ended questions that pertained to the administration and utility of both the CRT and also the CBM measures. The respondents were those in the building who were familiar with the CBM-M, CBM-R, CRT-M and CRT-R outcome measures, and had administered all on a regular basis. The respondents were the assistant to the principal/mathematician, the reading clinician, the student resource assistant, and the learning disabilities teacher.

QUESTION 1: WHAT ARE YOUR COMMENTS AND/OR CONCERS REGARDING THE READING AND MATHEMATICS CRTs?

This question addressed five different sub-areas: time spent in preparation, time spent in administration, accuracy of achievement levels, usefulness of results, and whether it can be assessed if the student has made progress/failure based on the CRT. The purpose of this question was to determine the teachers' opinions of the advantages and disadvantages of the CRT method.

QUESTION 2: WHAT ARE YOUR COMMENTS AND/OR CONCERNS REGARDING READING AND CBM?

This question addressed the same five different sub-areas: time spent in preparation, time spent in administration, accuracy of achievement levels, usefulness of results, and whether it can be assessed if the student has made progress/failure based on the results of the CBM measures. The purpose of this question was to determine the teachers' opinions of the advantages and disadvantages of the CBM method.

Procedures

As noted earlier, the population from which the sample was drawn was comprised of students and teachers from one elementary school in the Northwest suburbs of Chicago. Data collection commenced in September, 1993 and concluded in May, 1994.

Beginning in the fall of 1993, 20 students were randomly selected from each grade-level roster and were administered three randomly selected CBM-R probes. Each student was individually administered the three different reading passages by either the school psychologist, the school psychology intern, or the reading clinician. Each of three examiners were carefully trained by the methods advocated by Shinn (1989). Each subject was given 1-minute in which to respond to each probe. The examiner recorded the words per minute (WPM) for each of the three passages and then calculated the median score of all three passages for each subject.

In the fall of 1993, every student in the building was administered a CBM-M probe. The school psychologist, the school psychology intern, and the reading clinician each went into the different classrooms and administered the CBM-M probe to the entire class. The examiner then recorded the number of correct digits (CD) for each student.

It should be noted that a different mathematics and three different CBM reading probes were administered in the winter and in the spring. CBM norms were then developed for reading and mathematics by the school psychologist and the school psychology intern, utilizing the methods according to Shinn (1989). A graphic representation of the means, medians, and standard deviations was created for each grade level.

In April of 1994, every student in the building was administered the Cognitive Abilities Test (CogAT) according to the directions set forth by the test publishers. Test levels administered were 2, A, B, and C respectively for grades 2, 3, 4, and 5. The only students excluded from this test were those who were eligible for special education and their IEP excluded them from such tests. In the second grade, items were read one at a time by the test administrator and students chose answers which they marked in designated booklets. The test took three sessions to administer. In the third, fourth, and fifth grades, students completed the three subtests independently. Thirty minutes were provided to each subject to complete each of the three sections of the test.

In May of 1994, every student in the school was administered the CRT-R which corresponded to their grade level. Also in May, but on a different day from the CRT-R, every student was administered a grade-level appropriate form of the CRT-M.

Again, due to children moving and child absences, the total number of students participating in the study varied across grade levels.

Statistical Analyses

To test Null Hypotheses 1-4, the Macintosh-based program, Statview SE+ Graphics was used. In addition, series of correlation analysis procedures were used to analyze the data sets, examine comparisons, determine levels of significance, and to provide information to facilitate the interpretation of the findings. These correlation coefficients were systematically examined to determine the relationship between the two assessment methods (CBM and CRT) (Shaughnessy & Zechmeister, 1990). Descriptive procedures were used to determine whether differences arise in the identification of students at risk for academic failure when the identification is based on the CBM procedures versus the CRT measures.

CHAPTER FOUR

RESULTS

This chapter is divided into three subsections. The first section related to testing Null Hypotheses 1-4 focuses on the research questions pertaining to the examination of the relationships between curriculum-based measurement and the CRT measure (a published criterion-referenced achievement test). The second section describes the data set related to addressing the research questions involving whether the CBM measure targets the same or different students compared to the CRT procedures. In addition, an effort is made to determine whether the CBM outcome measures can identify students who are at risk for academic failure earlier in the academic school year when compared to the CRT procedures (i.e., Hypotheses 5-8). The third subsection provides a fine grained description of the results of the questionnaire that four teachers completed.

In order to determine the relationship between the two assessment methods (CRT and CBM), correlation coefficients were systematically examined and compared. All correlations were found to be positive, with the exception of the correlation between CogAT-NV and CBM-R (third grade) and the correlation between CBM-M and CRT-M (fourth grade). Six variables were included in each matrix: CBM Reading (CBM-R); CBM Mathematics (CBM-M); CRT Reading (CRT-R); CRT Mathematics (CRT-M); CogAT Verbal (CogAT-V); and CogAT Nonverbal (CogAT-NV). Means and standard deviations for grades 2, 3, 4, and 5 for the following variables are listed in Table 4.1: CBM-R; CBM-M; CRT-R; CRT-M; CogAT-V; and CogAT-NV.

Table 4.1

Means and Standard Deviations for Reading, Mathematics, and Ability Measures for Second, Third, Fourth, and Fifth Grades

	Grade							
	2		3		4		5	
Measure	M	SD	M	SD	M	SD	M	SD
CBM-M	20.0	8.7	12.9	6.8	30.1	15.0	35.5	13.9
CBM-R	114.8	53.0	111.7	32.9	121.7	31.6	132.7	41.5
CRT-M	47.9	7.0	45.8	7.4	50.1	3.2	51.5	3.4
CRT-R	92.6	14.2	75.6	11.1	85.1	8.3	85.6	7.5
CogAT-V	103.8	14.2	99.9	13.2	105.3	11.2	107.7	12.9
CogAT-NV	109.2	15.6	103.6	11.3	108.2	13.3	106.7	12.0

Results Related to Testing Null Hypothesis 1

As indicated in Table 4.2, for second-grade respondents, a significant correlation was found between the CRT-R and the CBM-R scores. In addition, a highly significant correlation was found between the CRT-M and the CBM-M measures. A strong significant correlation was also found between the CRT-M and the CRT-R measures and between the CBM-M and the CBM-R outcome measures.

Results indicated that there was a higher but nonsignificant correlation between the CogAT-V score and the CRT-R outcome measure than between the CogAT-V score and CBM-R outcome measure. However, a significant correlation between the CogAT-NV score and the CRT-M outcome measure was found. Conversely, there was no significant relationship found between the CogAT-NV score and the CBM-M outcome measure. Also, a significant correlation was found between the CogAT-NV score and the CRT-R outcome measure, but not between the CogAT-NV score and the CBM-R outcome measure. Table 4.1 contains the means and standard deviations for the following variables for the second-grade respondents: CBM-M; CBM-R; CRT-R; CRT-M; CogAT-V; and CogAT-NV.

Table 4.2

Intercorrelations Between Variables for Second-Grade Respondents (n=16)

Variable	1	2	3	4	5	6
1. CBM-R	--					
2. CBM-M	.704*	--				
3. CRT-M	.494*	.561*	--			
4. CRT-R	.620*	.466	.874*	--		
5. CogAT-V	.081	.318	.517*	.388	--	
6. CogAT-NV	.213	.398	.611*	.533*	.428	--

* $p < .05$

Results Related to Testing Null Hypothesis 2

Results of the correlation matrix for the third grade are presented in Table 4.3. A highly significant relationship was found between the CRT-M scores and the CBM-M outcome measures. However, a weaker, nonsignificant correlation was found between the CRT-R scores and the CBM-R outcome measures, A significant correlation was found between the CRT-R outcome measures, the CRT-M scores, and the CBM-M outcome measures. In addition, there was a significant correlation found between the CBM-M outcome measures and the CBM-R outcome measures.

An examination of the matrix indicated that the CogAT-V scores were highly related to both the CBM-R outcome measures and the CRT-R scores. In addition, the CogAT-V score was found to correlate highly with both the CBM-M outcome measures and the CRT-M scores. The CogAT-NV score was correlated highly with the CRT-M and the CRT-R scores, but was not significantly related to the CBM-M and CBM-R outcome measures. Furthermore, there was a moderately high significant correlation found between the CogAT-V scores and the CogAT-NV scores. Table 4.1 contains the means and standard deviations for the third-grade respondents for the following variables used in the study; CBM-M; CBM-R; CRT-M; CRT-M; CogAT-V; and CogAT-NV.

Table 4.3

Intercorrelations Between Variables for Third-Grade Respondents (n=14)

Variable	1	2	3	4	5	6
1. CBM-R	--					
2. CBM-M	.532*	--				
3. CRT-M	.237	.677*	--			
4. CRT-R	.343	.756*	.928*	--		
5. CogAT-V	.577*	.744*	.609*	.730*	--	
6. CogAT-NV	-.097	.346	.670*	.626*	.504*	--

* $p < .05$

Results Related to Testing Null Hypothesis 3

Results of the correlation matrix from the fourth-grade sample, as shown in Table 4.4, indicate that the correlation between the CBM-R scores and the CRT-R measures is nonsignificant. Likewise there was no significant correlation found between the CBM-M and the CRT-M outcome measures. In addition, no significant correlation was found between the CBM-R scores and the CBM-M outcome measures. However, a strong correlation was found between the CRT-M and the CRT-R outcome measures. Given these results, a significant relationship is clearly indicated between the CRT-M and the CBM-R measures.

The results appearing in the correlation matrix table also indicate that there was a strong, significant correlation between the CogAT-NV scores and the CBM-M scores. However, there was a low, nonsignificant correlation found between the CogAT-NV score and the CRT-M outcome measures. A significant correlation was found between the CogAT-NV scores and both the CRT-R and the CogAT-V measures. With regard to the CogAT-V score, a highly significant correlation was found with the CRT-R outcome measure. However, no significant correlation was found between the CogAT-V score and the CBM-R outcome measure. In addition, a significant correlation was found between the CogAT-V score and the CRT-M outcome measure, but not with the CBM-M outcome measure. Table 4.1 contains the means and standard deviations for the fourth-grade respondents for the following variables used in the study: CBM-M; CBM-R; CRT-R; CRT-M; CogAT-V; and CogAT-NV.

Table 4.4

Intercorrelations Between Variables for Fourth-Grade Respondents (n=15)

Variable	1	2	3	4	5	6
1. CBM-R	--					
2. CBM-M	.061	--				
3. CRT-M	.527*	-.041	--			
4. CRT-R	.280	.313	.602*	--		
5. CogAT-V	.399	.247	.496*	.720*	--	
6. CogAT-NV	.063	.646*	.392	.578*	.482*	--

* $p < .05$

Results Related to Testing Null Hypothesis 4

As shown in Table 4.5, an examination of the correlation matrix from the fifth-grade student sample reveals a highly significant correlation between the CRT-R and the CBM-R outcome measures. In addition, a significant correlation was found between the CRT-M and the CBM-M measures. However, no significant correlations were found between the CRT-R scores and either the CBM-M and the CRT-M outcome measures. Likewise, no significant correlation was found between the CBM-M and the CBM-R outcome measures.

Strong, significant correlations were also noted between the CogAT-NV and both the CBM-M and the CRT-M measures. In addition, the CogAT-NV scores correlated significantly with both the CBM-R and the CRT-R outcome measures. High, significant correlations were found between the CogAT-V score and both the CBM-R and the CRT-R measures. However, a significant correlation was found between the CogAT-V scores and the CRT-M scores, but not with the CogAT-V scores and the CBM-M scores. A significant correlation was found between the CogAT-V scores and CogAT-NV scores. Table 4.1 contains the means and standard deviations for the fifth-grade respondents for the following variables: CBM-M; CBM-R; CRT-M; CRT-R; CogAT-V; and CogAT-NV scores.

Table 4.5

Intercorrelations Between Variables for Fifth-Grade Respondents (n=14)

Variable	1	2	3	4	5	6
1. CBM-R	--					
2. CBM-M	.461	--				
3. CRT-M	.088	.601*	--			
4. CRT-R	.629*	.440	.418	--		
5. CogAT-V	.520*	.294	.581*	.710*	--	
6. CogAT-NV	.511*	.643*	.671*	.659*	.658*	--

* $p < .05$

Summary of Results Related to Testing Null Hypotheses 1-4

Results of the correlation matrices indicate that coefficients differ significantly among the different grade levels. However, it should be noted that there are some commonalties across the grade levels. For example, it appears that in three out of the four grade levels (i.e., second, third, and fourth) there is a high correlation between the CRT-R and the CRT-M measures. Likewise, two of the lower grade level subjects (i.e., second and third) displayed significant correlations between the CBM-R and the CBM-M measures, while the upper grade subjects did not (i.e., fourth and fifth).

Significant correlations between the CRT-M and the CBM-M scores were evident at the second, third, and fifth grade levels, but not evident at the fourth grade level. Significant correlations between the CBM-R and the CRT-R scores was only evident at the second and fifth grade levels, but not evident at the third and fourth grade levels.

At the third and fifth grade levels, there was a significant correlation found between the CogAT-V scores and both the CRT-R and the CBM-R measures. At the second grade level, there was no significant correlation found between the CogAT-V score, the CRT-R, or the CBM-R measures. An examination of the fourth-grade matrix reveals a significant correlation between the CogAT-V scores and the CRT-R measures, but not between the CogAT-V scores and the CBM-R measures.

At the second and third grade levels, a significant correlation was found between the CogAT-NV scores and the CRT-M measures, but not between the CogAT-NV scores and the CBM-M measures. However, at the fourth grade level, there was a significant correlation found between the CogAT-NV scores and the CBM-M measures, but not between the CogAT-NV scores and the CRT-M measures. In addition, at the fifth grade level the CogAT-NV scores correlated with both the CBM-M measures and the CRT-M measures.

Results Related to Qualitative Research Questions 1-4

Once again, it should be noted that students who were identified as being at risk for academic failure using the CRT outcome measures were only identified in the spring. However, since CBM outcome measures were administered three times in the school year, students could potentially be identified by the CBM measures in the fall, winter, and/or spring.

Shinn (1989) discussed the utilization of two different cutting score procedures to determine a student's special education eligibility. The first method is based on the discrepancy ratio, which is calculated by dividing the greater academic performance, typically, that of general education students, by the lesser performance, typically special education students. It is recommended that there are different discrepancy ratios for each grade level. The second method used to determine cutting scores is a percentile rank procedure, which permits a user to identify the same number of students as eligible, regardless of their grade level. Typically, for special education eligibility, a 10th percentile cutting score has been used.

Because the examiner in this study was attempting to identify students at risk for academic failure, and not special education eligibility, two different methods were utilized to determine which students would be identified in the fall, winter, and/or spring using the CBM measures. Again, the methods used in this study were designed to be less cautious with respect to the number of students identified, due to the fact that special education eligibility was not a major consideration here.

The first method identified those students who fell one standard deviation or more away from the mean (CBM-1SD) when compared to the locally generated norms. Again, with relationship to the locally generated norms, the second method used the bottom quartile (CBM-25) as the cutoff for students at risk of academic failure.

Grade 2. As shown in Table 4.6, utilizing the CBM-1SD cutoff score, the two second grade students that were identified using the CRT-R in the spring were also identified when the CBM-R was used in the fall, winter, and spring. However, two additional students were also identified by using the CBM-R measure in the spring.

When changing the CBM cutoff score to the 25th percentile, two additional student that were not identified by the CRT-R measure in the spring, were identified as having academic difficulties in reading (one in the fall and one in the winter). Another student that was identified by the CBM-R measure in the spring with the CBM-1SD, was identified by the CBM-25 earlier in the school year (in the fall and winter).

Table 4.6

Number of Students the Reading CBM Measure Identified of the Two Second-Grade Students Identified by the Reading CRT Measure

Fall		Winter		Spring	
Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM
1 Standard Deviation					
2	0	2	0	2	0
Bottom Quartile					
2	2	2	2	2	0

In the area of mathematics, one second grade student was identified by the CRT-M measure in the spring. The findings appearing in Table 4.7 indicate that when using the CBM-1SD cutoff score, that the same student was also identified by the CBM-M measure in both the winter and spring, However, three additional students were identified with the CBM-M measures (one in the fall, one in the winter and spring, and one in the spring).

Conversely, when the CBM-25 cutoff score was used, the one student who was identified by using the spring CRT-M procedure was identified even earlier through using the CBM-M measure in the fall. One additional student, who was not identified by the

CRT-M procedure, was identified by using the CBM-M measure in the fall and winter. Also, one additional student was identified by the CBM-M measure in the fall

Table 4.7

Number of Students the Mathematics CBM Measure Identified of the One Second-Grade Students Identified by the Mathematics CRT Measure

Fall		Winter		Spring	
Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM
1 Standard Deviation					
0	1	1	1	1	2
Bottom Quartile					
1	2	1	1	1	0

Grade 3. A total of four students were identified using the spring CRT-R measures in the fourth grade. The findings reported in Table 4.8 illustrate that, when the CBM-1SD cutoff score was utilized, none of the four students identified by the CRT-R measures were identified by the CBM-R outcome measures. Although, one additional student was identified using the CBM-R measure in the fall, winter, and spring.

However, when the cutoff score was set at the 25th percentile, two of the students that were identified by using the spring CRT-R measures, were picked up by the CBM-R procedures (one in the winter and one in both the winter and spring). In addition, one more student was identified as having difficulty in reading using the CBM-R outcome measures in the fall, winter, and spring.

Table 4.8

Number of Students the Reading CBM Measure Identified of the Four Third-Grade Students Identified by the Reading CRT Measure

	Fall		Winter		Spring	
	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM
1 Standard Deviation	0	1	0	1	0	1
Bottom Quartile	0	1	1	1	2	1

Four students were identified using the spring CRT-M measures. As shown in Table 4.9, using the CBM-1SD cutoff score, two of the four students identified by the CRT-R measure were also identified by CBM-R measure; one student in both the winter and the spring and one student in the spring. Additionally, one student was identified by the CBM-M measure in the spring.

However, when the CBM-25 cutoff score is utilized, all four students who were identified by the CRT-M measures were identified even earlier through using the CBM-M measures (two students in the fall, winter, and spring, one student in the fall and spring, and one student identified in the winter). Also with the CBM-25 cutoff score, three additional students were identified by CBM-M as having mathematics difficulties in the fall, one additional at-risk student was identified in the fall and spring, and one additional at-risk student was identified in the winter.

Table 4.9

Number of Students the Mathematics CBM Measure Identified of the Four Third-Grade Students Identified by the Mathematics CRT Measure

Fall		Winter		Spring	
Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM
1 Standard Deviation					
0	0	2	0	3	1
Bottom Quartile					
3	4	3	1	3	1

Grade 4. As shown in Table 4.10, no fourth grade students were identified by the spring CRT-R. However, when identifying students at risk for academic failure with the CBM-1SD cutting score, one student was identified through using the CBM-R measures in the winter and the spring.

When identifying those students who fall below the CBM-25 cutting score, one student who was identified with the CBM-1SD in the winter and the spring, was identified with the CBM-25 earlier in the fall. In addition, three students were identified in the fall, winter, and spring by using the CBM-R outcome measures.

Table 4.10

Number of Students the Reading CBM Measure Identified of the Zero Fourth-Grade Students Identified by the Reading CRT Measure

	Fall		Winter		Spring	
	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM
1 Standard Deviation	0	0	0	1	0	1
Bottom Quartile	0	4	0	3	0	3

The results reported in Table 4.11 indicate that no fourth grade student was identified by the spring CRT-M. When identifying those whose scores fell below the CBM-1SD cutting score, one student was identified in the fall and one in the winter and spring.

When the CBM-25 cutoff score was utilized, a total of five students were identified by using the CBM-M scores (two in the fall, one in the fall and winter, one in the fall, winter, and spring, and one in the winter and spring).

Table 4.11

Number of Students the Mathematics CBM Measure Identified of the Zero Fourth-Grade Students Identified by the Mathematics CRT Measure

	Fall		Winter		Spring	
	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM
1 Standard Deviation	0	1	0	1	0	1
Bottom Quartile	0	4	0	3	0	2

Fifth Grade. A total of three fifth grade students were identified by using the CRT-R measures as being at risk for academic failure. As shown in Table 4.12, by identifying students with the CBM-1SD cutoff score, one of the three at risk students identified by the spring CRT-R measure, was also identified by the CBM-R measure in the fall, winter, and spring. In addition, one at risk student who was identified using the CRT-R measures was also identified using the CBM-R measures in the spring. However, one student who was identified using the spring CRT-R measure was not identified by using the CBM-R score. Also, one additional student was identified in the fall and spring by using the CBM-R score, but not and the CRT-R score.

Using the CBM-25 cutoff score for students at risk for academic failure, revealed that one student who was identified by using the CRT-R score was identified even earlier in the fall using the CBM-R measure. In addition, two more at risk student were identified in the winter using the CBM-R score.

Table 4.12

Number of Students the Reading CBM Measure Identified of the Three Fifth-Grade Students Identified by the Reading CRT Measure

Fall		Winter		Spring	
Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM
1 Standard Deviation					
1	1	1	0	2	1
Bottom Quartile					
2	0	1	2	2	0

As shown in Table 4.13, no fifth grade at risk student was identified using the spring CRT-M measures. Using the CBM-M measures, with the cutoff score set at CBM-1SD, revealed that one student was identified as being at risk for academic difficulties in mathematics in the fall, one student in the winter, and three students in the spring. When using the CBM-25 cutoff score, two additional students were identified as being at risk for academic difficulties in mathematics in the winter, one in the fall and winter, and one additional in the spring.

Table 4.13

Number of Students the Mathematics CBM Measure Identified of the Zero Fifth-Grade Students Identified by the Mathematics CRT Measure

	Fall		Winter		Spring	
	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM	Also Identified by CBM	Identified Only by CBM
1 Standard Deviation						
	0	1	0	1	0	3
Bottom Quartile						
	0	1	0	3	0	1

Summary of Results Related to Qualitative Research Questions 1-4

Using the cutoff criteria of at least one standard deviation away from the mean, an examination of the data set revealed that 20 of the 59 subjects in the study were identified as having some academic difficulty in reading and/or mathematics as indicated by either the CBM and/or the CRT measures. The grade distribution for the number of students identified was as follows: in the second grade, five students were identified; in the third grade, six students were identified; in the fourth grade, two students were identified; and in the fifth grade, seven students were identified.

When utilizing the cutoff score at the twenty-fifth percentile, 33 of the 59 subjects were identified as having some academic difficulty in reading and/or mathematics as indicated by the CBM and/or the CRT measures. Thus, thirteen more students were identified when the twenty-fifth percentile was used as the cutoff score versus the use of one standard deviation away from the mean as cutoff score. The grade distribution for the number of students identified was as follows: 7; 10; 8; and 8; for grades two, three, four, and five, respectively.

Results of the Survey

An open-ended survey was developed by the investigator to examine the teachers' thoughts and feelings related to the acceptability and utility of two assessment methods (curriculum-based measurement and a criterion referenced test). The survey was administered to four teachers at Byrd Elementary School. Their responses were as follows.

Question 1: What are your comments and/or concerns regarding the reading and mathematics CRT. . .

• regarding the time spent preparing?

Three of the four respondents indicated that the time needed to prepare for the CRT was too long. Indeed, one respondent stated that the preparation takes too much valuable time away from children learning.

• regarding time spent to administer?

Three of the four respondents revealed that the amount of time spent administering the CRT was very long and many days were actually needed. However, one of the respondents stated that the actual time to administer was not an issue.

• accuracy of achievement levels?

Three of the four respondents indicated a lack of assurance that the instrument was accurately measuring what it was intended to measure. One respondent stated that with any paper and pencil task, many variables impact test performance, and, thus that might be one explanation as to why the scores may not be reflective of individual level and/or ability. One respondent did not respond to this item.

• usefulness of results?

All respondents indicated that the usefulness of results for immediate instructional change is questionable. However, three of the respondents stated that the results helped for making instructional changes for the following year.

• whether or not you can determine if the student has made progress/failure in your classroom based on the CRT scores?

One respondent indicated that growth is shown, but it is of little value since the CRT is only given at the end of the year. In addition, one respondent added that it is difficult to know whether a student has obtained mastery on all levels. Two of the respondents did not respond to this item.

Question 2: What are your comments and/or concerns regarding reading and mathematics CBM…

• regarding time spent preparing?

All four of the respondents indicated that the preparation time for CBM is very minimal and not of concern.

• regarding the time spent to administer?

All respondents stated that they were pleased that CBM is quick to administer. In fact one respondent revealed that even her students look forward to CBM and the students quickly get the materials ready for the teacher.

• accuracy of achievement levels?

Three of the respondents said that they thought that CBM levels were an accurate representation of their students' achievement levels. One respondent indicated that she still feels somewhat unsure of the correlation between reading fluency and reading comprehension.

• usefulness of results?

All four respondents had positive things to say regarding the usefulness of the CBM results. Two of respondents indicated that CBM procedures really shows the ups and downs of the students' progress. In addition, another respondent stated that the results were a quick and easy tool to make preliminary instructional groups. Another respondent stated that the results of CBM were practical.

• whether or not you think you can determine if the student has made progress/failure in your classroom based on CBM scores?

All four respondents indicated that progress/failure in students' achievement levels can definitely be seen using the CBM procedures. In fact, one respondent added that the students enjoyed seeing a visual graph of their progress and/or their failure.

CHAPTER FIVE

DISCUSSION

In this final chapter, a summary of the study is presented along with a discussion related to the testing of the eight hypotheses. Following the summary and discussion of results, recommendations for further research and implications for schools are presented.

This study was designed to investigate the accuracy of predicting students' achievement levels using two different measures: 1) locally normed curriculum-based measurement (CBM) procedures; and 2) group standardized testing norms (CRT). More specifically, the study was designed to examine the correlation between the locally generated CBM norms and the group standardized norms.

Furthermore, the study was also designed to examine the accuracy of whether or not the CBM procedures can predict students who are at risk for academic failure earlier in the school year than the CRT. An effort was made to show that the CBM procedures can predict those students who need additional academic support as well, if not better, than a standardized, criterion-referenced test (CRT), which is only administered once a year in the spring.

The study was conducted in one elementary school, which was part of a school district serving a Northwest suburban community of approximately 6000 students. This district includes a diverse population of minority and low income children of approximately 10%.

In the fall, winter, and spring of the 1993-94 school year, 20 students selected randomly from second, third, fourth, and fifth grade were administered CBM reading passages and also CBM mathematics probes, following the specifications made by Shinn (1989). Again, following the guidelines by Shinn, local norms were generated for the school. In addition, each student was administered the Cognitive Abilities Test (CogAT) in April and the Reading Criterion-Referenced Test (CRT-R) and the Mathematics Criterion-Referenced Test (CRT-M) in May.

Discussion Related to Null Hypothesis 1

As indicated in Chapter 4, the statistical analyses of the data set related to testing this null hypothesis showed a significant correlation between the CRT-R and the CBM-R measures for the second grade student sample. This correlation was used to estimate the degree to which the curriculum-based measures correlated with the basal mastery tests (CRT-R). A significant correlation was also found between the CRT-M and the CBM-M measures for the second grade sample. This correlation provided an estimate of the degree to which the curriculum-based measures correlated with the district CRT-M.

Another significant correlation was found between the CRT-R and the CRT-M measures. In addition, a significant correlation was found between the CBM-R and the CBM-M outcome measures. No significant correlation was found between the CogAT-V scores and either the CRT-R or the CBM-R measures. However, the CogAT-NV scores were found to be significantly correlated with the CRT-M measures, but not the CBM-M measures.

Discussion Related to Null Hypothesis 2

An examination of the results of the statistical analyses of the data set related to testing this hypothesis indicated that there was no significant correlation between the CBM-R and the CRT-R measures for the third grade student sample. However, there was a significant correlation found between the CBM-M and the CRT-M measures for the third grade students. As with the second grade student sample, significant correlations were found between the CRT-R and the CBM-R measures and between the CRT-M and the CBM-M measures.

A significant correlation was found between the CogAT-V scores and the CBM-R outcome measures. Likewise a significant correlation was found between the CogAT-V scores and the CRT-R measures. The CogAT-NV scores were not found to be significantly correlated with the CBM-M scores, but a significant correlation was found between the CogAT-NV scores and the CRT-M measures.

Discussion Related to Null Hypothesis 3

The statistical analyses related to testing this null hypothesis showed that there was no significance correlation between the CBM-R and the CRT-R measures for the fourth grade students. In addition, there was no significant correlation found between the CBM-M and the CRT-M measures for the fourth grade students.

A significant correlation was found between the CRT-R and the CRT-M measures. However, unlike the results from the second and third grade student samples, there was no significant correlation found between the CBM-R and the CBM-M outcome measures. However, there was a significant correlation found between the CogAT-NV scores and the CBM-M measures, but not with the CogAT-NV scores and the CRT-M measures. Conversely, there was a significant correlation found between the CogAT-V scores and the CRT-R measures, but not with the CBM-R measures.

Discussion Related to Null Hypothesis 4

The statistical analyses of the results related to testing this null hypothesis indicated that there was a highly significant correlation between the CBM-R and the CRT-R measures for the fifth grade students. In addition, a significant correlation was found between the CBM-M and the CRT-M measures for the fifth grade students. However, no significant correlations were found between the CRT-R and the CRT-M measures and between the CBM-R and the CBM-M outcome measures.

Results also indicated that there was a significant correlation between the CogAT-V scores and both the CBM-R and the CRT-R measures. Likewise, a significant correlation was found between the CogAT-NV scores and both the CBM-M and the CRT-M measures.

Summary of Discussion Related to Null Hypotheses 1-4

Although only significant results were found between the CRT-R and the CBM-R measures in two of the four grade levels used in the study, these results provide additional support for the criterion-related validity of curriculum-based reading measures as a predictor of global reading proficiency. These results are compatible with the findings of many others (Tindal, Shinn, Fuchs, Fuchs, Deno, & Germann, 1983; Fuchs, Tindal, Shinn, Fuchs, Deno, & Germann, 1983; Fuchs, Tindal, Fuchs, Shinn, Deno, & Germann, 1983) who reported that curriculum-based reading measures shared a great deal of variance with basal mastery tests that correlated highly with general measures of reading skills. It should be noted that the correlations from previous studies were higher than the corrrelations found here. However, 8 of the 14 studies reviewed by Marston (1989) that related to the validity of using CBM reading measures involved studies with students across several grades grouped together. Clustered together, these correlations using students in Grades 1-6 grouped together, do not appear to be very informative with respect to determining the validity correlations within a grade (Mehrens & Clarizio, 1993).

Perhaps one reason why in two of the grade levels no significant correlations were found between the two reading measures (CBM-R and CRT-R) is that the publishers of the CRT-R measures only provided teacher judgments as a way to assess validity, and thus, the CRT-R measures may lack validity. It is possible that another reason related to why no significant correlations were found in two of the grade levels could be that the CRT-R measures and the CBM measures were influenced by a lack of overlap of reading material between the CRT measures and the Harcourt, Brace, & Jovanovich curriculum used for CBM measures. In fact, both Pany (1978) and Shapiro and Derr (1987) found biased curriculum content in individually administered achievement tests.

Indeed, another plausible explanation for the discrepancy of significant findings across grade levels could be that the magnitude changes in primary grades due to possibly individual differences in decoding skills (Jewell & Jenkins, 1993). Again another possible explanation and a limitation of the study was lack of control over the type of instruction used in the classroom and the amount and the extent of individualized preparation of the students for the tests.

Although only significant results were found between the CRT-M and the CBM-M measures in three of the four grade levels used in the study, these results provide additional support for the criterion-related validity of curriculum-based mathematics measures as a predictor of global mathematics proficiency. These results confirm other work (Skiba, Magnusson, Marston, & Erikson, 1986) in which curriculum-based measures correlated moderately with district CRT basic mathematics concepts. However, in all four studies reviewed by Marston (1989), students were grouped by multiple grade levels. In addition, for those studies in which the focus was on the district CRT as the criterion, the median coefficient was found to be .34 (Mehrens & Clarizio, 1993).

Analogous explanations posited for the inconsistency in correlations found for the reading measures, exist with the mathematics measures. Indeed, there was no technical adequacy information available for the CRT-M measures. Thus, one could speculate that the CRT-M measures were not reliable and/or valid. This possibility could have a significant impact on the reliability and/or validity of the CRT-M with the CBM-M measures. Again, I question whether or not the CRT-M measures had any overlap of curriculum content with the current curriculum used in the school. In addition, both the CBM and the CRT measures were found to vary across grades depending on the focus of skills.

In three grade levels a significant correlation was found between the CRT-R and the CRT-M measures. I could speculate that this may be due to the CRT-M measure assessing more than just mathematics. Likewise, a significant correlation was found between the CBM-R and the CBM-M outcome measures in the second and third grades, but not with the fourth and fifth grades. It should be noted that the reason for this correlation in two of the four grade levels could not be due to the CBM-M assessing more than just mathematics because the CBM-M measure is strictly mathematics computation and involves no reading.

In an attempt to determine concurrent validity, the Cognitive Abilities Test was used in this study. The CogAT-V was found to be significantly correlated with the CRT-R measure for the third, fourth, and fifth grade student samples. At the second grade level, neither the CRT-R or the CBM-R measures correlated with the CogAT-V. The CBM-R measure was found to be significantly correlated with the CogAT-V for the third and fifth grade student samples. It was expected that the CogAT-V would not correlate with either reading measure (CBM-R and CRT-R) at the second grade level since the CogAT-V at that level is not specifically measuring reading. In fact the CogAT-V for second graders is read to them in order to control for individual differences in reading.

Again, when attempts were made to assess concurrent validity between the CogAT-NV and the CRT-M and the CBM-M measures, the CogAT-NV was found to be significantly correlated with the CRT-M for the second, third, and fifth grade student samples. The CogAT-NV was significantly correlated with the CBM-M measure for the fourth and fifth grade student samples.

Discussion Related to Qualitative Research Questions 1-4

Second Grade: Reading. With the CBM-1SD cutoff score, both students who were identified by the CRT-R measure were also identified by the CBM-R measure in the fall, winter, and spring. In addition, use of the CBM-R measure picked up two additional students who were at risk for academic failure. When the cutoff score changes to CBM-25, four additional students who were not identified by the CRT-R measure, were identified by the CBM-R measure. Furthermore, two of these four students who were identified by the CBM-R measure, were identified within the first two months of the school year.

Second Grade: Mathematics. One student who was identified by the CRT-M measure was also identified in the winter by the CBM-M measure with the cutoff score at CBM-1SD. In addition, with the same cutoff score, three more students were identified earlier in the school year with the CBM-M measure. When the cutoff score was changed to CBM-25, the one student who was identified by the CRT-M measure was also identified even earlier in the fall with the CBM-M measure. Also, with the CBM-25 cutoff score, one more student in the winter was identified as having academic difficulties in mathematics.

Third Grade: Reading. The CRT-R measure identified four third-grade students from the third-grade student sample. However, when using the CBM-1SD cutoff score, none of the four students were identified by the CBM-R measure. Perhaps these four students were not identified by the CBM-R measure because the cutoff score was too high.

When the cutoff score was changed to CBM-25, two of the four students identified by the CRT-R measure were identified by the CBM-R measure. I could speculate that the reason the other two students were not identified by the CBM-R measure with the larger cutoff score, is that the two students in question could have been very good oral readers but had poor comprehension skills. Indeed, when examining their CRT-R subtest scores, both students showed a significant weakness in the area of reading comprehension. A study completed by Shinn, Good, Knutson, Tilly, & Collins (1992) indicated that for the third grade students there does not appear to be a distinction between decoding and comprehension constructs. However, the CBM reading measures where students read aloud from third-grade basal readers correlated only moderately with inferential and literal comprehension measures ($r = .71$ and $r = .72$, respectively), while the correlation is higher when using oral reading fluency as a index of reading decoding ($r = .88/.90$).

Third Grade: Mathematics. The CRT-M measure identified four third-grade students from the third-grade student sample. When utilizing the CBM-1SD, three of the four students identified by the CRT-M measure were identified by the CBM-M measure. In fact, when the cutoff score changed to CBM-25, all four of the students identified by the CRT-M measure were also identified by the CBM-M measure. Furthermore, when the CBM-M measure was used, all four students were identified even earlier in the school year than when using the CRT-M (e.g., three of the four students were identified within the second month of school).

Fourth Grade: Reading. No fourth-grade students from the fourth-grade student sample were identified by the CRT-R measure. With the CBM-1SD cutoff score, one at risk student was identified by the CBM-R measure in the winter and the spring. When the cutoff score was changed to CBM-25, a total of four students were identified by the CBM-R outcome measure in the fall, winter, and spring. Thus, using the larger cutoff score, allowed for more students to be identified and, thus, be eligible to receive additional instructional support.

Fourth Grade: Mathematics. Again, no fourth-grade students from the fourth-grade student sample were identified by the CRT-M measure. With the CBM-1SD cutoff score, two students were identified by the CBM-M measure. However, the number of students identified by the CBM-M measure increased to six students when the CBM-25 cutoff score was utilized. Furthermore, four of the six students identified by the CBM-M measure were identified within the second month of school.

Fifth Grade: Reading. The CRT-R measure identified three fifth-grade students from the fifth-grade student sample. When the CBM-1SD cutoff score was used, two of the three students identified by the CRT-R measure were also identified by the CBM-R measure. Perhaps a reason why the other student who was identified by the CRT-R measure was not identified by the CBM-R measure, was that the cutoff score was too high. However, when the cutoff score was changed to CBM-25, that same student was not identified by the CBM-R outcome measure. I could speculate that a reason for this is due to the fact that this student had received a great deal of additional resource services for the past several years. These additional resource services included test taking strategies and test accommodations (i.e., tests administered individually and tests read aloud to students) that all regular education students did not receive. In addition to receiving this additional instructional support, the student also received CBM progress monitoring weekly to help facilitate the student's oral reading fluency and reading comprehension.

Fifth Grade: Mathematics. No fifth-grade students from the fifth-grade student sample were identified by the CRT-M measure. However, five students were identified by the CBM-M measure when the cutoff score was CBM-1SD. When the cutoff score changed to CBM-25, those five students were identified even earlier in the school year with the CBM-M measure. In addition, with the cutoff score of CBM-25, a total of eight students were identified by the CBM-M outcome measure.

Conclusions

Based upon the results of this study, the CBM reading passages developed from the Harcourt, Brace, & Jovanovich reading series, appear to have adequate criterion-related validity when compared to the CRT measures, although the strength of these findings varies somewhat within grade levels. The CBM mathematics probes appeared to have adequate criterion related validity when compared to the CRT-M measures. Again, the strength of these findings varies somewhat within grade levels. Taken together, these findings contribute to the technical adequacy of the CBM procedures and provide continued support to those who choose to use the CBM procedures as an additional measure for screening and instructional decision making.

In this study, two different cutoff criteria (CBM-1SD and CBM-25) were used to determine which students would be identified as being at risk for academic failure in the fall, winter, and/or spring using the CBM measures. Because this study was designed in an effort to identify students at risk for academic failure, and not special education eligibility, the results of the study support the view that the CBM-25 cutoff score was the preferred method. The CBM-25 cutoff score allowed more at risk students to be identified earlier in the school year. Consequently, these identified students could begin receiving additional instructional support sooner than those not identified. In addition, the number of at risk students identified using the CBM-25 cutoff score was larger than the number of at risk students identified using the CBM-1SD cutoff score. In sum, more students were identified as being at risk using the CBM-25 cutoff score compared to the

CBM-1SD cutoff score and were given instructional support in a hopeful manner before academic problems developed.

The results of the survey conducted in this study indicated that all respondents thought the CRT measures took a large amount of time to prepare and administer, and thus, took time away from teaching students. However, all respondents indicated that the preparation time and administration time were very brief with the CBM measures. A majority of the respondents reported that they were unsure whether or not the CRT measures were accurately measuring what they were intended to measure. On the other hand, a majority of the respondents reported that they thought the CBM measures were an accurate representation of the students' achievement levels. With the CRT measure, all respondents indicated that the utility of the data for instructional planning were questionable. Conversely, all respondents indicated that the CBM measures were useful and provided the teacher with a wealth of data. In summary, the results of this study, along with those reported by others (Eckert, Shapiro, & Lutz, 1995; Wilson, Schendel, & Ulman, 1992), support the use of curriculum-based measurement as a more acceptable method of assessment than published, standardized achievement tests.

Currently, Admiral Byrd School does not have any standards for determining which students are at risk for academic failure and would, thus, require resource support in order to facilitate their success. Byrd School, with a student population of approximately 400, is fortunate to have a large amount of resource services available to students. These resource services consist of a full-time mathematician, a full-time reading clinician, a full-time student resource assistant, and a part-time teacher for at risk students.

In addition, this study was designed in an effort to provide evidence that the CBM measures allow individual students to receive better educational services to meet their unique needs in a timely fashion. In order to do so, the CBM data sets must yield quantitative and qualitative descriptions of student performance in order for educators to determine when instruction needs to be adapted and enhanced (Fuchs & Deno, 1994).

Data from this study and from many others, provide support for using the CBM measures as a prereferral intervention in order to identify at risk learners and provide sufficient educational support services. Indeed, prereferral interventions exemplify an education practice that addresses the needs of at-risk learners in general education (Bahr, 1994). Also, prerefferal intervention can reduce referrals for special education and increase the accuracy of placement rates for children who are referred for being at-risk for academic failure (Ysseldyke, Pianta, Christenson, Wang, & Algozzine, 1983).

Taking the findings from others (Shinn, 1989; Shinn & Habedank, 1992; Tindal, 1992; Fuchs & Deno, 1991; Fuchs & Fuchs, 1991) along with the findings of the study reported here, CBM measures do appear to be an effective and efficient method for screening and monitoring progress of students throughout the academic school year. The CBM procedures appear to be a viable, data-driven method for determining which students need additional instructional supports. In addition, CBM can be utilized to monitor the effectiveness of instruction, and to determine when instructional change is appropriate and/or necessary.

In summary, although curriculum-based measurement provides documentation of the effectiveness of interventions and determines whether students are making adequate progress, the purpose of this study was not to promote the use of the CBM measures as an isolated set of tests. Instead, the CBM measures should be used in conjunction with many current educational assessment practices. When CBM is used with other sources of information, CBM provides a useful, defensible way to gather information about students' needs.

Limitations of the Study

Results from this study were based on data from one elementary school. The investigator assumed that this sample can be used to represent problem identification based on achievement testing for all schools. However, it is recognized that the results may not generalize well to other schools within the school district, or other districts. In addition, results are based on a relatively small sample size of 59 students from only four different grade levels.

Finally, it should be noted that another limitation of the study is that some of the subjects who were involved in this study were simultaneously receiving additional academic support outside of the classroom. The decision as to which subjects would receive the additional support, and for how long, was not based solely on the data reported here, but rather on idiosyncratic teacher referrals.

Suggestions for Further Research

Given the results of this study, additional research appears warranted. Topics for further investigation include: (a) support for the addition of an alternative CBM reading measure; and (b) support for the addition of an alternative CBM mathematics measure.

Although results of this study provide support for the CBM reading passages as an indicator of reading skills, including reading decoding and comprehension, the use of the cloze, retell, and maze methods may be useful as instructional methods or diagnostic strategies for determining directions for instruction (Fuchs & Fuchs, 1992). The cloze method leaves the first sentence of a passage intact, but thereafter, every *n*th word is omitted and replaced with a blank. The subject is then required to restore meaningful deletions. This method appeared to load more highly on reading comprehension compared to the oral reading fluency scores. This finding is important to note because despite a series of published validation studies, questions about whether or not oral reading fluency measures reading comprehension continue to persist (Shinn, Good, Knutson, Tilly, & Collins, 1992). The retell method involves subjects reading passages and then retelling in their own words what occurred in the passage, without referring back to the text. One disadvantage of this method is that scoring can be a difficult and time-consuming process. It appears that both the cloze and retell methods may be useful as instructional methods or diagnostic strategies for determining directions for instruction. However, when investigating criterion validity for these two methods (i.e., cloze and retell) they were found to be technically unsatisfactory for feasibly and accurately monitoring student growth across time.

Another method, the reading maze procedure, can be used to monitor reading and was investigated by Fuchs, Fuchs, Hamlett, & Ferguson (1992). The format of this method consisted of the first sentence of a passage remaining intact. Thereafter, every *n*th word is deleted, and replaced with three choices. The subject, who is timed, then is asked to select an alternative that meaningfully replaces each blank. This method appeared to be useful for monitoring student reading growth. In addition, the criterion validity was found to be strong and the technical features were similar to that of oral reading fluency. Likewise, the maze method was found to be an acceptable measure of reading, which indexes decoding, fluency, and comprehension (Fuchs & Fuchs, 1992).

Perhaps using one or any of these three methods (retell, cloze, and/or maze) in combination with the CBM oral reading fluency measure could increase the validity of the measures. In addition, more pertinent information could be gathered regarding an individual's strengths and weaknesses.

In the area of mathematics, another method that could be used in combination with the CBM mathematics computation measure, is the CBM mathematics concepts and application methods developed by Fuchs, Fuchs, Hamlett, Thompson, Roberts, Kubek, and Stecker (1994). This method requires students to perform both grade-level mathematics problem solving activities and mathematics computations, thus expanding CBM mathematics to incorporate the broader mathematics curriculum. For example, the

CBM mathematics concepts and applications addresses number concepts, counting, applied computation, geometry, measurement, charts, graphs, money, and problem solving. Fuchs et. al. (1994) reported that the CBM mathematics concepts and applications system can be used as a tool to help the teacher design more effective programs in the area of mathematics, concepts, applications, and problem solving. In addition, results from their study support the technical adequacy of the CBM concepts and applications system. Thus, the information derived from this procedure has the possibility to be accurate and meaningful for educators.

APPENDIX A

SAMPLE CURRICULUM-BASED MEASUREMENT READING PASSAGES

FOR SECOND, THIRD, FOURTH, AND FIFTH GRADE

Grade 2

One day Abigail's family moved away, across wide rivers and over a rock-hard trail. The quilt went too. It was not stuffed into the trunks. It kept Abigail and her sisters warm from the wild winds. It kept them warm from the rain and the cold nights.

Abigail's father built a new house in the woods. He built Abigail a new bed. He made her a new wooded horse, too. When Abigail's father was finished, everyone said, "Welcome home."

Abigail felt sad. They had a new house, a new horse, and a new bed. Everything was new, except the quilt. So Abigail's mother rocked her as mothers do. Then she tucked her in, and Abigail felt at home again under the quilt.

One day when the quilt was very old and very loved, Abigail folded it carefully and put it in the attic. Many years passed. Everyone forgot the quilt was in the attic.

Grade 3

A thrill of excitement ran through the children. "We'll keep this a secret, okay?" said Wilford. "If some smart grown-ups hear what's down in that hole, they'll buy this land in a hurry. They'll make a lot of money by charging people three dollars a ticket to see those cave drawings!"

The children nodded. There was a fortune in it!

"I can rent this land," said Wilford, "but I need a little more money. Then I can dig an opening to give people a better way to get into the lower cave."

"I knew you'd ask us for money," said Rocky Graham. He was a member of the Tigers, a club for tough older boys.

"Get lost, kid," said Wilford. To all others he said, "I'm going to let each and every one of you buy a piece of this business for five dollars. We'll all make a fortune."

"How do we know that those walls have cave drawings on them?" asked Benny Breslin.

"After I found those cave drawings, I went home and got my camera," said Wilford. "I took pictures with a flash."

He passed out three photographs. The first was of a wooly rhinoceros. The second was of cave people attacking a dinosaur. The third was of a charging mammoth.

Grade 4

Suddenly the rain began to slacken, and I walked around the house. I had never been so wet in my life. Now that it was over I was cold, too, and tired. I looked up at the tree and there didn't seem to be any point in climbing back up. In just a few hours everyone would know what I had done anyway. I went up on the porch and rang the doorbell.

It was Aunt Millie in her cotton robe who turned on the porch light and peered out through the side windows at me.

I must have been an awful sight, for she flung open the door at once and drew me in.

"What are you doing out there? What are you doing?"

"Who is it?" Uncle Fred asked as he came into the hall.

"It's Tom," Aunt Millie said.

They both turned and looked at me, waiting for an explanation. I cleared my throat and said, "Uncle Fred and Aunt Millie, I am awfully sorry but I have let the baby fox out of the rabbit hutch." I sounded very stiff and formal, and I thought the voice was a terrible thing to have to depend on, because I really did want them to know that I was sorry, and I didn't sound it the least bit. I know how much Uncle Fred had looked forward to the hunt and how important getting rid of the fox was to Aunt Millie, and I hated for them to be disappointed now.

Grade 5

One morning Tom sat in the glazing room, working mechanically, trying not to think about Meg. His eye fell on the slate with the pictures he had drawn for her. Taking up a large, unfinished plate, he dipped his brush in cobalt blue and began to copy his picture onto the clay surface. Father would not mind his taking the time if it pleased Meg. She must get better! He would put a whole story on the plate and then glaze it for her.

There should be people on the plate. Meg would want people in the picture. Tom drew a man in a boat, with a long pole to move the boat along. There was room for three more little figures on the bridge. He drew a woman, copying the robes of a Chinese lady he had seen on a teacup, then a man, then another man. Was this last man chasing the first couple? Yes, perhaps he was.

The center of the plate was full now. The different parts of the picture made a most pleasing design. Tom began to decorate the edge of the plate, imitating the patterns that ran around the rims of the Chinese ware. This careful work took the rest of the day. Finally it was done. Then he glazed the plate and set it with the other ware to be baked in the kiln on the following day.

That night no one slept until dawn. The doctor stayed near the little girl all night. The first light of morning was beginning to break when he came into the kitchen where the anxious family huddled.

“Her fever has broken,” he said. “she should get well now.”

APPENDIX B

SAMPLE CURRICULUM-BASED MEASUREMENT MATHEMATICS PROBES FOR SECOND, THIRD, FOURTH AND FIFTH GRADE

Grade 2

3 x 5	7 + 6	13 -8	11 -7
742 -542	58 + 25	8 +4	13 -9
4 3 +5	3 x 5	16 -7	7 + 9
13 +20	8 + 12	777 + 115	9 +7
15 -9	636 - 264	8 + 4	6 + 3

Digits Correct: ________________

Grade 3

22 x3	25 x3	601 - 486	4) 18
49 +61	87 x5	6,009 -2,324	50 -36
485 +608	605 +327	641 -397	2) 52
85 -37	512 372 + 429	22 +48	24 x9
12 x9	3) 23	39 x5	604 - 196

Digits Correct: ____________

Grade 4

35,721 +9,845	4 x4	12) 32	2 x2
32 x13	851 - 285	40÷8=	16 x6
46 x2	8) 65	6 x8	7) 846
2) 612	3) 21	9) 47	22) 92
46) 73	25 x41	601 - 388	30÷6=

Digits Correct: ______________________

Grade 5

27,677 + 19,281	9) 5,570	6,117 + 5,089	667 x900
5) 986	58) 4,682	22) 129	811 x546
59) 8,892	7) 1,617	4) 172	366 x42
84) 416	4,942 -1,988	3,770 -1,308	5) 573
37) 232	700 - 186	6) 3,550	192 x346

Digits Correct: ________________

APPENDIX C

TEACHER SURVEY

Please take your time and answer the following questions relating to the achievement assessment of the children at Byrd School.

1. What are your comments and/or concerns regarding the reading and mathematics CRTs…

   a. regarding time spent preparing?

   b. regarding the time spent to administer?

   c. accuracy of achievement levels?

   d. usefulness of results?

   e. whether or not you can determine if the student has made progress/failure in your classroom based on the CRT scores?

2. What are your comments and/or concerns regarding the reading and mathematics CBM…

   a. regarding time spent preparing?

   b. regarding the time spent to administer?

c. accuracy of achievement levels?

d. usefulness of results?

e. whether or not you can determine if the student has made progress/failure in your classroom based on the CBM scores?

References

Allen, D. (1989). Periodic and annual reviews and decisions to terminate special education services. In M. R. Shinn (Ed.), Curriculum-based measurement: Assessing special children (pp. 184-203). New York: Guilford.

Anania, J. (1982). The effects of quality of instruction on the cognitive and affective learning of students. (Doctoral dissertation, University of Chicago, 1981). Dissertation Abstracts International, 42, 4269A.

Anania, J. (1983). The influence of instructional conditions on student learning and achievement. Evaluation in Education: An International Review Series, 7(1), 1-92.

Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan.

Armbruster, B. B., Stevens, R. G., & Rosenshine, B. (1977). Analyzing content coverage and emphasis: A study of three curricula and two tests (Technical Representative N26). Urbana, IL: University of Illinois, Center for Study of Reading.

Bahr, M. W. (1994). The status and impact of prereferral intervention: "We need a better way to determine success". Psychology in the Schools, 31, 309-318.

Bain, S. K., & Garlock J. W. (1992). Cross-validation of criterion-related validity for CBM reading passages. Diagnostique, 17(3), 202-208.

Baker, E. L., O'Neil, H. F., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. American Psychologist, 48(12), 1210-1218.

Bloom, B. S. (1995). The search for methods of instruction. In A. C. Ornstein & L. S. Behar (Eds.), Contemporary issues in curriculum (pp. 208 - 225). Boston: Allyn and Bacon.

Burke, A. J. (1984). Students' potential for learning contrasted under tutorial and group approaches to instruction. (Doctoral dissertation, University of Chicago, 1983). Dissertation Abstracts International, 44, 2025A.

Cizek, Gregory J. (1995). The big picture in assessment and who ought to have it. Phi Delta Kappan, 77(3), 246-250.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.

Deno, S. L. (1986). Formative evaluation of individual student progress: A new role for school psychologists. School Psychology Review, 15, 358-374.

Deno, S. L. (1987). Curriculum-based measurement. Teaching Exceptional Children, 41-47.

Deno, S. L. (1992). The nature and development of curriculum-based measurement. Preventing School Failure, 36(2), 5-10.

Deno, S. L., Marston, D., Shinn, M., & Tindal, G. (1983). Oral reading fluency: A simple datum for scaling reading disability. Topics in Learning and Learning Disability, 2, 53-59.

Deno, S. L., Marston, D., & Tindal, G. (1986). Direct and frequent curriculum-based measurement: An alternative for educational decision making. Special Services in the Schools, 2(2-3), 5-27.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. Exceptional Children, 18, 16-26.

Derr-Minneci, T. F., & Shapiro, E. S. (1992). Validating curriculum-based measurement in reading from a behavioral perspective. School Psychology Quarterly, 7(1), 2-16.

Eckert, T. L., Shapiro, E. S., & Lutz, J. G. (1995). Teachers' ratings of the acceptability of curriculum-based assessment methods. School Psychology Review, 24(3), 497-511.

Elliott, S. N., & Bretzing, B. H. (1980). Using and updating local norms. Psychology in the Schools, 17, 196-201.

Freeman, D. J., Kuhs, T. M., Porter, A. C., Floden, R. E., Schmidt, W. H., & Schwille, J. R. (1983). Do textbooks and test define a national curriculum in elementary school mathematics? Elementary School Journal, 83, 501-513.

Fuchs, L. S., Butterworth, J. R., & Fuchs, D. (1989). Effects of ongoing curriculum-based measurement on student awareness of goals and progress. Education and Treatment of Children, 12(1), 63-72.

Fuchs, L. & Deno, S. (1981) The relationship between curriculum-based mastery measures and standardized achievement tests in reading (Research Report No. 57) Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Fuchs, L. S., & Deno, S. L. (1991). Effects of curriculum within curriculum-based measurement. Exceptional Children, 58(3), 232-243.

Fuchs, L. S., Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. Exceptional Children, 57(5), 488-500.

Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? Exceptional Children, 61(1), 15-24.

Fuchs, L. S. & Fuchs, D. (1986). Effects of systematic formative evaluation on student achievement: A meta-analysis. Exceptional Children, 53, 199-208.

Fuchs, L. S., & Fuchs, D. (1991). Curriculum-based measurements current applications and future directions. Preventing School Failure, 35(3), 6-11.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. School Psychology Review, 21(1), 45-58.

Fuchs, L. S., Fuchs, D., & Deno, S. L. (1982). Reliability and validity of curriculum-based informal reading inventories. Reading Research Quarterly, 18, 6-26.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1990). Curriculum-based measurement: A standardized, long-term goal approach to monitoring student progress. Academic Therapy, 25(5), 615-632.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1991). Effects of expert system advice within curriculum-based measurement on teacher planning and student achievement in spelling. School Psychology Review, 20, 49-66.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement, using a reading maze task. Exceptional Children, 58(5), 436-451.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. Exceptional Children, 60(6), 518-538.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1990). The role of skills analysis in curriculum-based measurement in math. School Psychology Review, 19,(1), 6-22.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Thompson, A., Roberts, P. H., Kubek, P, & Stecker, P. M. (1994). Technical features of a mathematics concepts and applications curriculum-based measurement system. Diagnostique, 19(4), 23-49.

Fuchs, L. S., Fuchs, D., & Maxwell S. (1988). The validity of informal reading comprehension measures. Remedial and Special Education, 9(2), 20-28.

Fuchs, L. S., Fuchs, D., & Warren, L. M. (1982). Special education practice in evaluating student progress toward goals (Research Report No. 81). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Fuchs, L. S., Hamlett, C. L., Fuchs, D., Stecker, P. M., & Ferguson, C. (1988). Conducting curriculum-based measurement with computerized data collection: Effects on efficiency and teacher satisfaction. Journal of Special Education, 9,(2), 73-86.

Fuchs, L. S., Wesson, C., Tindal, G., Mirkin, P. K., & Deno, S. L. (1982). Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures (Research Report No. 64). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Gathercoal, Paul. (1995). Principles of assessment. The Clearing House, 69(1), 59-62.

Germann, G., & Tindal, G. (1985). An application of curriculum-based assessment: The use of direct and repeated measures. Exceptional Children, 52(3), 244-265.

Gickling, E. E., & Havertape, J. (1981). Curriculum-based assessment (CBA). Minneapolis, MN: National School Psychology Inservice Training Network.

Glass, G. V. & Hopkins, K. D. (1984). Statistical methods in education and psychology (2nd ed.). Boston: Allyn and Bacon.

Good, T. L., & Brophy, F. E. (1984). Looking in classrooms (3rd ed.). New York: Harper and Row.

Good III, R. H., & Shinn, M. R. (1990). Forecasting accuracy of slope estimates for reading curriculum-based measurement: Empirical evidence. Behavioral Assessment, 12(2), 179-193.

Jamentz, K. (1994). Making sure that assessment improves performance. Educational Leadership, 51(3), 55-58.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. (1979). Measuring pupil progress toward the least restrictive environment. Learning Disability Quarterly, 2, 81-92.

Jenkins, J. R. & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. Exceptional Children, 59(5), 421-432.

Jitendra, A. K., & Kameenui, E. J. (1993). Dynamic assessment as a compensatory assessment approach: A description and analysis. Remedial and Special Education, 14(5), 6-18.

Knutson, N., & Shinn, M.R. (1991). Curriculum-based measurement: Conceptual underpinnings and integration into problem-solving assessment. Journal of School Psychology, 371-393.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Eds.), Curriculum-based measurement: Assessing special children (pp. 18-78). New York, NY: The Guilford Press.

Marston, D., & Deno, S. L. (1982). Implementation of direct and repeated measurement in the school setting (Research Report No. 106). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Marston, D., Deno, S. L., Mirkin, P. K., Lowry, L., Sindelar, P., & Jenkins, J. (1981). An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study. Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Marston, D., Deno, S. L., & Tindal, G. (1984). Eligibility for learning disabilities services: A direct and repeated measurement approach. Exceptional Children, 50, 554-555.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. Exceptional Children, 52, 266-276.

Marston, D. & Magnusson, D. (1988). Curriculum-based assessment: District-level implementation. In J. L. Graden, J. E. Zins, and M. C. Curtis (Eds.). Alternative educational delivery systems: Enhancing instructional options for all students (pp. 137-172). Washington, DC: National Associations of School Psychologists.

Marston, D., Mirkin, P., & Deno, S. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. Journal of Special Education, 18(2), 109-117.

McCurdy, B. L., & Shapiro, E. S. (1992). A comparison of teacher-, peer-, and self-monitoring with curriculum-based measurement in reading among students with learning disabilities. Journal of Special Education, 26(2), 162-180.

Mehrens, W. A., & Clarizio, H. F. (1993). Curriculum-based measurement: Conceptual and psychometric considerations. Psychology in the Schools, 30(3), 241-254.

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35(11), 1012-1027.

Mirkin, P., Deno, S. L., Tindal, G., & Kuehnle, K. (1982). Frequency of measurement and data utilization as factors in standardized behavioral assessment of academic skill. Journal of Behavioral Assessment, 4(4), 361-370.

Oakland, T., & Matuszek, D. (1977). Using tests in a non-discriminatory fashion. In T. Oakland (Ed.), Psychological and educational assessment of minority children (pp. 52-69). New York: Bruner-Mazil.

Ornstein, A. C., & Hunkins, F. (1993). Curriculum foundations, principles, and theory (2nd ed.). Boston: Allyn and Bacon.

Potter, M.. L., & Wamre, H. M. (1990). Curriculum-based measurement and developmental reading models: Opportunities for cross-validation. Exceptional Children, 57(9), 16-25.

Riverside 2000: Technical summary 1. (1994). Chicago: Houghton Mifflin.

Roberts, M. L., Turco, T. L., & Shapiro, E. S. (1991). Differential effects of fixed instructional ratios on students' progress in reading. Journal of Psychoeducational Assessment, 9(4), 308-318.

Ryan, J. M. & Miyasaka, J. R. (1995). Current practices in testing and assessment: What is driving the changes? NASSP Bulletin, 79(573), 1-11.

Salmon-Cox, L. (1981). Teacher and standardized achievement test: What's really happening? Phi Delta Kappan, 62, 631-634.

Shapiro, E. S. (1990). An integrated model for curriculum-based assessment. School Psychology Review, 19, 331-400.

Shapiro, E. S., & Derr, T. F. (1987). An examination of overlap between reading curricula and standardized achievement tests. The Journal of Special Education, 21, 59-67.

Shapiro, E. S., & Eckert, T. L. (1994). Acceptability of curriculum-based assessment among school psychologists. Journal of School Psychology, 32, 167-184.

Shaughnessy, J.J. & Zechmeister, E. B. (1990). Research methods in psychology (2nd ed.). Boston: McGraw-Hill Publishing Company.

Shinn, M. R. (1988). Development of curriculum-based local norms for use in special education decision-making. School Psychology Review, 17(1), 61-80.

Shinn, M. R. (1988). Local norms for use in special education decision-making. School Psychology Review, 17(1), 61-80.

Shinn, M. R. (Eds.). (1989). Curriculum-based measurement: Assessing special children. New York, NY: Guilford Press.

Shinn, M. R., Good III, R. H., Knutson, N., Tilly III, W. D., & Collins, V.L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. School Psychology Review, 21(3), 459-479.

Shinn, M. R., & Habedank, L. (1992). Curriculum-based measurement in special education problem identification and certification decisions. Preventing School Failure, 36(2), 11-15.

Shinn, M. R., Habedank, L., Rodden-Nord, K., & Knutson, N. (1993). Using curriculum-based measurement to identify potential candidates for reintegration into general education. The Journal of Special Education, 27(2), 202-221.

Shinn, M. R., & Hubbard, D. D. (1992). Curriculum-based measurement and problem-solving assessment: Basic procedures and outcomes. Focus on Exceptional Children, 24(5), 1-20.

Shinn, M. R. & McConnell, S. (1994). Improving general education instruction: Relevance to school psychologists. School Psychology Review, 23(3), 351-371.

Shinn, M. R., Nolet, V., & Knutson, N. (1990). Best practices in curriculum-based measurement. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology-II (pp. 275-286). Washington, DC: National Association of School Psychologists.

Shinn, M. R., Tindal, G., & Spira, D. (1987). Special education referrals as an index of teacher tolerance: Are teachers imperfect tests? Exceptional Children, 54, 132-140.

Shinn, M. R., Tindal, G., & Stein, S. (1988). Curriculum-based assessment and the identification of mildly handicapped students: A research review. Professional School Psychology, 3, 69-85.

Shinn, M. R., Ysseldyke, J. E., Deno, S. L., & Tindal, G. A. (1986). A comparison of differences between students labeled learning disabled and low achieving on measures of classroom performance. Journal of Learning Disabilities, 19, 545-552.

Skiba, R., Magnusson, D., Marston, D., & Erikson, K. (1986). The assessment of mathematics performance in special education: Achievement tests, proficiency tests, or formative evaluation? Minneapolis: Special Services, Minneapolis Public Schools.

Slavin, R. E. (1980). Cooperative learning. Review of Education Research, 50, 314-342.

Thurlow, M. L., & Ysseldyke, J. E. (1982). Instructional planning: Information collected by school psychologists vs. information considered useful by teachers. Journal of School psychology, 20, 3-10.

Tindal, G. (1992). Evaluating instructional programs using curriculum-based measurement. Preventing School Failure, 36(2), 39-42.

Tindal, G., Germann, G., & Deno, S. L. (1983). Descriptive research on the Pine County norms: A compilation of findings (Research Report No. 132). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Tindal, G., Marston, D., & Deno, S. L. (1983). The reliability of direct and repeated measurement (Research Report No. 109). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Wesson, C. L. (1992). Using curriculum-based measurement to create instructional groups. Preventing School Failure, 36(2), 17-20.

White, O. R. & Haring, N. G. (1980). Exceptional Teaching (2nd ed.). Columbus, OH: Charles E. Merrill.

Wilson, M. S., Schendel, J. M., & Ulman, J. E. (1992). Curriculum-based measures, teachers' ratings, and group achievement scores: Alternative screening measures. Journal of School Psychology, 30(1), 59-76.

Ysseldyke, J. E., Pianta, B., Christenson, S., Wang, J., & Algozzine, B. (1983). Analysis of prereferral interventions. Psychology in the Schools, 20, 184-190.

VITA

The author, Lara E. Druckman, was born on October 19, 1969 in Skokie, Illinois.

Ms. Druckman obtained the Bachelor of Science degree in psychology from Bradley University and the Masters of Education degree in school psychology from Loyola University Chicago.

In June, 1993, she successfully completed an internship in school psychology with Community Consolidated School District #59, where she currently practices as a full-time school psychologist.

The dissertation submitted by Lara E. Druckman has been read and approved by the following committee:

Ronald R. Morgan, Ph.D., Director
Associate Professor, Curriculum, Instruction, and Educational Psychology
Loyola University Chicago

Joy Rogers, Ph.D.
Professor, Curriculum, Instruction, and Educational Psychology
Loyola University Chicago

Jack Kavanagh, Ph.D.
Professor, Counseling Psychology
Loyola University Chicago

Pamela M. Stecker, Ph.D.
Assistant Professor, School of Education
Clemson University

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the committee with reference to content and form.

The dissertation is, therefore, accepted in partial fulfillment of the requirements for the degree of Ph.D.

11/1/96
Date

Ronald R. Morgan
Director's Signature