5-1-2002

Must Curriculum Based Measurement Reading Probes Be Curriculum Based?

Nicole Riley-Heller
University of Nebraska at Omaha

Must Curriculum Based Measurement Reading Probes Be Curriculum Based?

An Ed.S. Field Project

Presented to the

Department of School Psychology

and the

Faculty of the Graduate College

University of Nebraska

In Partial Fulfillment

of the Requirements for the Degree

Educational Specialist

University of Nebraska at Omaha

by

Nicole Riley-Heller

May, 2002

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

ED.S. FIELD PROJECT ACCEPTANCE

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

Committee

Chairperson

Date 7/8/02

Contents

Must Curriculum Based Measurement Reading Probes Be Curriculum Based?

Nicole Riley-Heller, Ed.S.

University of Nebraska, 2002

Advisor: Lisa Kelly-Vance, Ph.D.

Curriculum-based measurement (CBM) has evolved as a reliable and valid method for measuring and monitoring student performance in basic academic skills. While the efficacy of CBM for assessing reading skills is not in question, there is some confusion regarding the type of CBM reading passages (called probes) that should be used. Specifically, it is unclear whether or not there is a difference between CBM probes derived directly from the instructional curriculum and generic probes derived from a pre-printed source. The current study compares the utility of two types of CBM reading probe materials. One probe type is derived from curriculum-dependent passages taken directly from a phonics-based reading instruction curriculum and the second from the Tests of Reading Fluency (TORF), a set of generic (curriculum-independent) passages. Both types of probes were administered to 13 second grade students twice weekly for 5 weeks. The median number of words read per minute for each probe type were compared. No significant differences were found between the two probes' in terms of ability to measure current levels of progress and performance over time. Implications of the study, outcomes for school psychology and potential for further research are discussed.

Must Curriculum Based Measurement Reading Probes Be Curriculum Based?

With the reauthorization of the Individuals with Disabilities Education Act (IDEA), Public Law 105-17 at the forefront of the School Psychology profession it was not surprising that a great deal of interest has been generated around alternatives to traditional (i.e. norm and criterion referenced tests) assessment methods. The potential changes in IDEA and the revisions in state laws that may undoubtedly result, lent strong support to the need for school psychologists to build additional measurement alternatives into their assessment skills repertoire.

School psychologists have traditionally used several methods to evaluate student achievement within the school setting. Most often, these methods have been categorized as either norm-referenced or criterion-referenced materials. Despite longstanding acceptance and widespread currency, these traditional methods of assessing academic performance have been subjected to scrutiny and criticism (Deno, 1985; Elliot & Fuchs, 1997; Knutson & Shinn, 1991; Shapiro, 1996). Both norm-referenced and criterion-referenced measures have provided indirect evaluation of skills by assessing students on a sample of items taken from expected age or grade level performance. Unfortunately, because these measures may not have been sensitive to small gains in student progress, it is possible that they have not been accurate indicators of whether intervention methods were effective (Shapiro, 1996). In addition, the items selected from these measures may have related only marginally to the actual content of the student's curriculum and the

resulting assessment may not have tested what was actually being taught (Shinn, Good, & Stein, 1989).

A common method of evaluating academic skills within school settings has involved the administration of commercially produced, norm-referenced, standardized tests. According to Shapiro, 1996, norm-referenced tests have been used extensively in schools to support decisions regarding eligibility for special education by providing a standardized comparison across peers of similar age and/or grade. Scores on standardized instruments have aided in identifying the degree to which an assessed student has deviated from his or her peers. Despite this asset, norm-referenced tests have been, by their nature, insensitive to small changes in student behavior (Fuchs, Fuchs, Hamelett, Walz & Germann, 1993). Furthermore, norm-referenced instruments were not designed to contribute to the development of intervention procedures (Elliot & Fuchs, 1997), and the information generated from these measures may not have reflected what was actually being taught in the classroom (Shinn, Knutson, Good, Tilly, & Collins, 1992).

Within school settings, another frequently used approach for assessing academic skills has been criterion-referenced tests. These assessment instruments may have provided some advantages over norm-referenced measures in that they were referenced to specific domains of behavior and had the potential to provide intrasubject rather than intersubject comparisons (Shaprio, 1996). In addition, criterion-referenced tests offered information on sub skills within broader areas, enabling the identification of specific skills to target in the development of intervention strategies. Despite the advantages

associated with criterion-referenced tests, they have not adequately addressed questions pertaining to educational classification, monitoring of student progress, and the development of intervention strategies. Furthermore, failure to demonstrate relevance to curriculum (i.e., lack of correspondence between items tested and actual curriculum), inability to provide information regarding short-term academic progress, and difficulty in the selection of subskills to assess may have restricted the effectiveness of these instruments and limited their application in the school setting (Shapiro, 1996).

Limitations associated with norm-referenced and criterion referenced measures created a demand within schools for the development and subsequent application of alternative assessment methods that were more directly relevant to learning in the classroom. In response to this demand, curriculum-based measurement (CBM) was developed and evolved as an alternative assessment measure that effectively addressed the limitations of other methods due to a direct link to the school curriculum (Deno, 1985).

CBM was a model of curriculum-based assessment developed by Stanley Deno and Phyllis Mirken (Marston, 1996). Specifically, CBM was a set of brief ongoing assessments with the goals of measuring individual student changes in short-term progress as well as evaluating student progress over time. CBM has been effectively utilized to measure student progress in the areas of reading, mathematics, computation, spelling, and written expression (Deno, 1985; Shinn 1989). CBM assessment has been demonstrated to be an objective, reliable, and valid standardized procedure directly

related to a student's academic skills within the actual school curriculum (Deno, 1985; Shinn 1989). Furthermore, the positive psychometric properties of CBM have received substantial support in the literature (Deno, Mirkin, & Chiang, 1982; Deno, 1985; Fuchs, Fuchs, & Maxwell, 1988; Fuchs & Shinn, 1989; Marston and Magnusson, 1985; Marston, 1989).

Due to frequent recurrent measurement, the data generated by CBM have been sensitive to short-term changes in student progress. In addition to providing frequent standardized repeated measurement, CBM measures were found to closely relate to the content of instructional curriculum (Fuchs & Deno, 1991; Howell, Fox, & Morehead, 1993). Consequently, the sensitivity of CBM measures to small changes in progress and the connection of CBM to what has been taught enabled an accurate indication of instructional effectiveness and aided in the development of appropriate instructional interventions (Deno, 1992; Fuchs & Deno, 1991; Shinn, 1989).

An advantage associated with CBM data has been the potential for the development of local norms. In fact, Deno (1985) identified the contributions of local norms to the decision making process as a defining feature of CBM. Marston (1989) maintained developing local norms with CBM screening and eligibility procedures provided critical and helpful information for school psychologists in that the data operationalized the expectations of the mainstream environment. In addition, Shinn (1989) maintained local norms could be used to support special education determinations. In an extension of this position, Knutson and Shinn (1991) indicated CBM has been used

effectively to determine academic difficulties and to ascertain the extent of those difficulties by operationalizing the discrepancy between expected student achievement and actual student achievement.

Marston (1989) indicated three levels within the schools at which local norms could be developed with CBM: 1) the classroom, 2) the entire school, and 3) at the school district level. Habedank Stewart and Kaminski (2002) identified several reasons why local school norms were desirable at all the levels indicated by Marston. One reason was that local norms enabled comparison of specific pertinent information about a student's academic performance relative to the performance of other local students with comparable educational backgrounds. Another reason identified by Habedank Stewart and Kaminski was that local norms have been useful in connecting CBM normative data with educational decision making by comparing students skills with local goals and standards. A final reason and benefit associated with the development of local norms was the potential for school psychologists to develop a database to aid in consultation and decision making.

The original intent in the development of CBM assessment measures was to derive materials directly from the instructional curriculum. According to Deno (1985), CBM was based on the premise that assessment and decision making were curriculum referenced. Essentially, a student's performance on a test should have indicated the student's level of competence in the local school curriculum. Hargis (1987) maintained deriving CBM materials directly from the curriculum insured test validity and the

relevance of instruction that resulted from such measurement. Similarly, Knutson and Shinn (1991) asserted, CBM ensured content validity of a measure due to accessing assessment materials directly from the curriculum of instruction.

Despite support for using instructional curriculum in the development of CBM assessment materials, a concern has been the lack of consistency within curriculum, especially if resulting CBM data were to be used for the development of local norms. According to Powell-Smith and Bradley-Klug (2001) this concern has been especially relevant to the area of reading, wherein variability has been recurrent in the types of curricular materials used. Fuchs and Deno (1994) indicated that there have been substantial differences in the difficulty levels of curriculum within the same reading series or even the same book.

Within the schools, variability has also occurred with the employment of multiple reading series' within the same setting. Given the popularity of inclusion (i.e. instruction of regular education and special education students within the same classroom) and its widespread application in the schools, it has not been uncommon for several reading series' to be used within a single classroom. In addition, within the same school building, numerous reading curricula have frequently been employed. Thus, the potential for a profusion of varied curriculum at a district level has been great.

Another concern identified by Habedank Stewart and Kaminski (2002) associated with deriving CBM reading assessment materials directly from the actual curriculum was the potential for advocating for the application of a specific published curriculum within

a classroom, building, and/or district; thus ensuring that all students were instructed with the same set of curriculum materials. Instructing all students in one curriculum, however, was in direct conflict with the basic tenet identified by Shinn, Good and Stein (1989), that student's who receive special education services are entitled to instructional programs that best fit their individual needs. Shinn, Good, and Stein maintained that "specialized instructional programs will result in greater achievement than might be obtained otherwise" (p.356). Consequently, within the school setting diverse curricula have been developed and applied to effectively instruct based on individual student strengths and areas of concern.

Diverse curricula have been beneficial in that specialized instruction was developed based on individual student needs. However, comparing all students to their peers with local norms has also been important in that norms provided a single accountability system for all children, both special and regular education. Local norms enabled comparisons of a student's academic performance to the performance of others, thus aiding in determining individual student needs. As Habedank Stewart and Kaminski (2002) indicated, local norms have been useful in connecting CBM normative data with educational decision making. For example, there may have been opportunities to utilize local norming data to aid in making determinations about placing a student in a specific reading curriculum based on their progress relative to peers.

Given the potential for variability in curriculum and concerns regarding promotion of a specific reading curriculum, it was desirable to have a constant source for

CBM reading assessment materials. Perhaps most important was ensuring consistency by controlling difficulty level while also maintaining a connection to educational outcomes. Fuchs and Deno (1994) concluded the use of CBM materials not derived from the curriculum, but controlled for difficulty, may have resulted in a reduction of measurement error. Thus, a common source related to, but not directly derived from the instructional curriculum that was controlled for difficulty level may have provided the best source for the development and subsequent application of CBM reading materials.

Consistency in CBM reading materials would have also facilitated the development of local norms. Habedank Stewart and Kaminski (2002) maintained that local norms "should reflect important educational outcomes" (p.740). They further indicated that it was not necessary for CBM reading materials to be derived directly from the school's curriculum, especially when the generated data were intended to be used for the development of local norms. Fuchs and Deno (1992) concluded deriving assessment materials directly from the curriculum of instruction was not a critical element of instructionally useful measurement. Concurrently, Powell-Smith and Bradley-Klug (2001) have posited that generic materials (which assessed the same skills required by the school's reading curriculum) were as valid as curriculum derived materials in measuring a student's current levels of performance and progress over time.

Due to the proposal that a link to the curriculum was not an essential feature in the development of reading passages, generic pre-printed reading passages that were controlled for difficulty level have been developed in lieu of curriculum derived probes

as a constant source for use in monitoring reading progress. Preliminary examination suggests generic (i.e. not derived directly from the instructional curriculum) CBM reading passages (called probes) may have been a viable alternative to reading probes derived directly from the curriculum (Powell-Smith & Bradley-Klug, 2001). Shapiro (1996) purported that the objective of evaluating reading with CBM was to find out where, in a graded set of materials, a student's reading skills fell. Shapiro further asserted that, when generic reading probes have been used, evaluators "should not be concerned about the apparent lack of a link to the curriculum of instruction, given that the reading material used for assessment is likely to be comparable to the instructional material" (p.111).

Several factors supported additional examination of the utility of CBM reading probes derived directly from instructional curriculum versus the utility of generic CBM reading probes derived from a pre-printed source was merited. First, while Powell-Smith and Bradley-Klug (2001) offered preliminary evidence that generic probes were comparable to curriculum derived probes, the examination was not without limitations. Second, the CBM process would have been less labor intensive and considerably faster if para-educators, teachers and school psychologists were not responsible for developing numerous reading probes directly from student's reading curriculum. Third, different reading curricula could have been implemented at all levels without interfering with the development of a single accountability system for all students. Fourth, potential for the promotion of a specific reading curriculum may be lessened. Finally, if an entire class,

building or school district used the same generic probes, it would have enabled and encouraged the development of local norms.

In summary, while the efficacy of CBM to assess reading skills has not been in question, there has been some confusion regarding the type (generic vs. curriculum derived) of CBM reading probe that should be used to effectively measure student progress. Due to the potential benefits associated with the application of generic CBM reading materials it may have been advantageous to ascertain whether or not there was a difference between probe types in capacity to measure current levels of student reading performance and student reading progress over time. The critical question therefore, was whether CBM reading probes had to be derived directly from the instructional reading curriculum to be effective in monitoring current levels of performance and progress over time (slope) or whether generic probes were just as effective.

Review of the Literature

Curriculum Derived Assessment

Knutson and Shinn (1991) presented a case study in an attempt to exemplify how CBM may be implemented into problem solving and decision making within a school setting. Knutson and Shinn identified three core principles that they maintained defined CBM. The principles that characterized CBM were that measures were individually referenced, norm-referenced and curriculum-referenced. According to Knutson and Shinn, the most important feature of CBM measures was that they were considered individually referenced in that they were intended to measure student's current rates of

academic progress over time. The current rates of progress served as a measure to be compared to previous progress in an effort to determine whether or not student achievement was commensurate with expected rate of progress in the curriculum. Knutson and Shinn maintained another defining feature of CBM materials was they were norm-referenced, in that local norms could be derived from the data generated with CBM. In addition, they asserted, determining expectations for students was accomplished by using CBM measures of local normative performance. Finally, they maintained an essential element of CBM was that materials were curriculum-referenced, meaning, "assessment materials are drawn directly from the curriculum in which students receive instruction" (p. 373).

While the position of Knutson and Shinn intuitively makes sense, there may have been potential for conflict between ensuring CBM measures were both curriculum-referenced and norm-referenced. Of consideration would have been varied readability levels within the same curriculum or different curriculum used within the same norm sample. According to Fuchs and Deno (1994) the potential for variability in curriculum has been great. The literature indicates that curriculum variability has not been conducive to the development of local norms (Fuchs & Deno, 1994; Habedank Stewart & Kaminski, 2002). Thus, it may not have been desirable to define CBM as both curriculum-references and norm-referenced when curriculum-referenced has been defined as meaning derived directly from the actual instructional curriculum.

Implications of Curriculum

The necessity of deriving assessment material directly from the curriculum was examined by Fuchs and Deno (1994) in their review of literature on the similarities between testing materials and curricula and through comparison of measurement validity versus instructional utility in assessing student performance with materials derived from instructional curriculum versus materials that were curriculum independent. In addition, Fuchs and Deno discussed an advantage and some disadvantages associated with using actual instructional material to monitor student progress.

Fuchs and Deno (1994) submitted an advantage of using instructional material for progress monitoring was the teacher's perception of participation in the process. Specifically, they posited that teachers were familiar with their class curriculum and, as a result, may have derived confidence in, or comfort with, assessment tools derived directly from that curriculum. Subsequently, this confidence may have translated into a vested interest and increased teacher utilization of this type of assessment. Thus, Fuchs & Deno, 1994 maintained "a major advantage of sampling performance using curricular materials appears to be face validity" (p. 18).

Despite this apparent advantage, several potential disadvantages of using the curriculum in which a student was being instructed were discussed. Fuchs and Deno (1994) submitted that because of wide ranges of readability within curriculum derived reading materials; those materials did not ensure control for difficulty level possibly resulting in measurement error. Another disadvantage was that students may have

experienced a practice effect due to having been previously exposed to the curricular reading materials from which probes were derived. The final disadvantage noted by Fuchs and Deno associated with deriving assessment materials directly from the instructional curriculum stemmed from concern that some reading curricula were highly controlled for vocabulary. The degree of assessment success was contingent on the specific reading material that the assessment was derived from limiting the generalizability to other reading texts.

Fuchs and Deno (1994) concluded it was not critical to use instructional curriculum in the development of instructionally useful measurements. They argued that if sampling material from curriculum did not ensure measurement validity or instructional utility and the disadvantages of using the curriculum outweighed the potential advantages, an examination should have been made regarding what features would have insured instructional utility of measurement. Accordingly, Fuchs and Deno derived three critical elements of instructionally valid measurement. They maintained that material chosen for assessment must have (a) enabled repeated testing over time on material of comparable difficulty, (b) possessed the capacity to accurately indicate and assess outcomes of instruction, and (c) provided quantitative and qualitative feedback regarding student performance.

Curriculum Independent Assessment

In support of the contentions of Fuchs and Deno (1994), Powell-Smith and Bradley-Klug (2001) conducted a study of 36 second-grade students from the Pacific

Northwest. They investigated the differences between generic versus curricular reading probes in monitoring students' oral reading fluency over a 5-week period. Powell-Smith and Bradley-Klug reported a statistically significant difference between measured basal curriculum derived probes words per minute (wpm) and generic Test of Reading Fluency (TORF) wpm (e.g., on average students read more wpm in TORF probes). No statistically significant differences were reported in the two probe types ability to measure progress over time. The rate of progress (slope metric) did not differ significantly when measured by either generic or curriculum derived reading probes.

While the Powell-Smith and Bradley-Klug (2001) study offered some preliminary insight on the curriculum issue, there were limitations. One or more of these limitations may have accounted for the statistically significant difference the authors found in the level of performance between the two kinds of reading probes (students read on average about 12 words per minute more on the generic reading probes). Limitations included application of two different methods for placement of students into monitoring materials. The generic screening probes and selection criteria were used for the generic probes whereas survey level assessment procedures were used for the basal (curriculum) probes. Additionally, the difference between wpm scores may have been the result of failure to counterbalance presentation of probe types. Specifically, basal/curriculum-dependent probes were always administered first followed by TORF probes. This presentation may have resulted in order effects. Another limitation of this study was that, rather than administering 3 reading probes per probe type and using the median number of words

read per minute from the 3 probes as Shinn (1989) intended, Powell-Smith and Bradley-Klug used only 1 probe per probe type, which may have limited application to actual CBM practice. Finally, in analyzing their data, Powell-Smith and Bradley-Klug combined all of the generic probe scores to yield a single mean, combined all of the curricular probe scores to yield a single mean, and then compared these 2 means. This decreased the power of their analysis and increased the probability of a Type II error.

Readability Formulas

According to Powell-Smith and Bradley-Klug (2001), when research has been conducted in the area of reading various techniques are frequently employed to assess the readability of the reading passages by applying one or more "readability formulas." Powell-Smith and Bradley-Klug further maintain that the intent of readability assessment has been to ascertain the level of difficulty associated with reading passages and to aid in making decisions regarding the impact of the determined difficulty level on students' reading performance. Essentially, readability programs were accessed to determine a student's instructional level (Wheeler & Sherman, 1983).

A concern with readability formulas has centered on how readability determinations were derived. Klare (1988) identified a few problems associated with the application of readability formulas. Central to these concerns was that one readability formula applied to a specific reading passage might have yielded different results than another readability formula applied to the same passage due to variations in criteria used

in the calculations. Essentially, individual readability formulas only accounted for their individual definitions of what constitutes level of difficulty in reading passages.

An additional concern with the use of readability formulas noted by Powell-Smith and Bradley-Klug (2001) was the amount of time associated with calculating the formulas. In their analysis of two types of readability formulas, Powell-Smith and Bradley-Klug concluded the two types were not consistent in their estimates of passage difficulty across reading curricula. They further concluded that readability data might have been useful in reading research; however, the amount of time allocated to calculating readability when developing CBM reading probes had the potential to be a hindrance for practitioners in the school setting (i.e. calculating readability consumed too much time).

Despite some concern about readability formulas, they have still been used extensively in the area of reading research (Powell-Smith & Bradley-Klug, 2001). The practice of assessing readability of passages has also been applied during the development of CBM reading probes. According to Shapiro (1996), when developing CBM reading materials, "...it is important to use passages that are carefully controlled for grade-level readability..." (p. 42).

Current computer technology may have accounted for some of the concerns associated with the application of readability formulas. Many word processing programs (e.g., Word Perfect, Microsoft Word) offered built-in readability formulas that incorporated multiple readability formula methods. By including multiple methods

several definitions of what constitutes level of difficulty in reading passages are accounted for. Shinn (1989) maintained that computer readability programs that provided a comparison of multiple methods were the most desirable and accurate.

The readability program from Microsoft Word not only provided a multiple method comparison, it also allowed for ready assessment of the readability of passages. A readability formula could have been automatically calculated by simply choosing the option via computer settings. To ensure standardization of presentation and consistent with the CBM procedures outlined by Shinn (1989), when developing CBM reading assessment materials, probes must be retyped in a common format to minimize effects of print type. Thus, when practitioners have been retyping CBM reading probes with a computer program such as Microsoft Word, they have had the option to automatically assess readability without having to spend any additional time with calculations.

Current Study

The purpose of the present study was to extend the research of Powell-Smith and Bradley-Klug (2001) by determining whether or not there was a difference in the utility of curriculum derived versus generic CBM reading probes. Consistent with Powell-Smith and Bradley-Klug, the current study was an attempt to address whether or not the probe types varied significantly in capacity to measure current levels of performance and whether probe types differed significantly in ability to measure progress over time or slope.

The CBM reading probe materials utilized for the current study were derived from two different sources. One source was the phonics-based reading series (Foresman, 2000) being used for reading instruction in the school where the study took place. This source was termed curriculum dependent probes (CDP). CBM literature documents strong reliability and validity for CBM probes derived directly from instructional reading curricula. Test-retest and alternate-form reliabilities on CBM curriculum derived reading measures sampled from 18 regular education and 15 special education students over a five-week period ranged from .73 to .91, with the majority of coefficients above .80 (Deno, et al., 1982). In a replication of Deno, et al. (1982), Marston and Magnusson (1985) found correlations ranging from .80 and .90 between CBM curriculum derived materials and subtests of the Stanford Achievement Test, the Science Research Associates, (SRA) Achievement series (Naslund, Thorpe, & Lefever, 1978) and Ginn 720 reading Series (Clymer & Fenn, 1979). In later examination of CBM curriculum derived reading materials, Marston (1996) found test-retest reliability correlations ranged from .82 to .97. Additionally, Marston's analysis of alternate-form reliabilities ranged from .84 to .96.

The other source for CBM reading materials was the Test of Reading Fluency (TORF), developed by Deno, Deno, Marston, and Marston (1987). The TORF was a set of standardized reading probes designed to identify student's current areas of reading difficulties. TORF probes were also developed with the intent of measuring individual student reading skills. Powell-Smith and Bradley-Klug (2001) identified TORF probes as

generic measures of reading because they were not derived from or directly affiliated with any specific reading curriculum. Little reliability data pertaining specifically to the TORF was provided in the tests manual. Median reliability coefficients of alternate forms of TORF screening passages ranged from .89 to .97 (Deno et al., 1987). Powell-Smith and Bradley-Klug (2001) maintained that despite the lack of validity studies specific to TORF reading probes, the probes likely retained psychometric properties similar to curriculum derived CBM reading probes. Consistent with CBM procedures delineated by Shinn (1989), every reading probe was examined via computer for grade level readability.

Every Tuesday and Thursday for 5 weeks, thirteen 2nd grade students were administered a set of 3 CDP and a set of 3 TORF probes. Different probes were utilized for all presentation sessions (e.g. the same student was never administered the same probe more than once) and all probes were counterbalanced prior to presentation. The number of correct words per minute (wpm) was recorded for each of the six 1 minute administrations. Each day, after all 3 reading probes from both sets were administered; the median number of wpm was identified from the 3 CDP scores and the 3 TORF scores. This resulted in 2 median wpm scores per day per participant, a total of 10 median wpm scores for each probe type over the data collection period. The dependent variable was median wpm, assessed both in terms of current level of performance and progress over time. The independent variables were the probe types, CDP or TORF. Based on the results of preliminary studies (Fuchs & Deno, 1994; and Powell-Smith & Bradley-Klug,

2001), two hypotheses were formulated. The first hypothesis was that there would not be a significant difference between CDP and TORF probes in effectiveness for monitoring current levels of performance in oral reading. The second hypothesis was that there would not be a significant difference between the 2 probe types in utility for measuring progress over time. A multiple regression with a dummy variable controlling for specific child effects was used to test the two hypotheses.

Method

Participants and Setting

Thirteen second grade students (7 males and 6 females) from 3 classrooms at a public school in the rural Midwest were participants for this study. Students from the second grade were selected because they have been identified as "emerging" readers (Snow, Burns, & Griffin, 1998), and therefore, have the most potential for growth over the data collection period. The students were 40% Hispanic and 60% Caucasian. Mean age for the students was 8 years 7 months. Students ranged in age from 8 years 1 month to 9 years 9 months. Participants were required to be able to speak and read in English (due to the reading probes having been written in English). All participants in the study were eligible for free or reduced lunch.

Students who participated in this study were identified by their teachers as low performers in reading, but not receiving special education services. Lower level readers were defined as those students performing in the lowest reading group in their classrooms

or as those students identified by teachers as being among the five least proficient readers in the class (Powell-Smith & Bradley-Klug, 2001). All participants were currently receiving reading instruction from a phonics-based reading series.

Procedure

Data Collection. Data were collected every Tuesday and Thursday for 5 weeks on student oral reading fluency in both CDP and TORF probes using curriculum-based measurement procedures (Deno, 1985). Five weeks enabled the collection of 10 data points; Shinn (1989) suggested that a minimum of 10 points be collected to obtain a reliable slope estimate. Data collection took place at the student's school during regular school hours. Probes were derived from passages of approximately 250 to 300 words in length in accordance with procedures outlined by Shinn (1989). Passages written in poetic or dramatic format were excluded. All CDP probes were derived from the Scott Foresman (2000) reading series. All CDP and TORF probes were retyped via Microsoft Word in a common format to minimize effects of print type. Every probe was also examined via Microsoft Word for readability in an effort to obtain the level of difficulty of the reading passage. Reading probes were packaged for each individual student and each packet contained both a set of 3 CDP and 3 TORF reading probes. Examiner and student copies of both types of measures were included in the packet. The only difference between the examiner and student copies was that the examiner copies contained a cumulative count of the number of words per line on the far right hand side of the page to

facilitate scoring. Different CDP and TORF probes were used for all sessions throughout the data collection period.

Administration and Scoring. Reading achievement was assessed by having students read aloud CDP and TORF probes estimated to be at an instructional level for students. Instructional level was defined as one grade level above where the student was currently mastering reading. In order for probes to be considered at an instructional level, students were required to read between 20 and 60 words correct per minute (Powell-Smith & Bradley-Klug, 2001). According to Fuchs and Shinn (1989), assessment materials should be selected for their potential for maximizing sensitivity of measurement, thus instructional level probes provided the advantage of ensuring that the difficulty level of probes remained relatively constant throughout the study period which may have resulted in increased sensitivity of measurement and aided in preventing floor and ceiling effects. Instructional level placement was determined by application of CBM assessment in both probe types in successively higher or lower levels until an appropriate grade level (where the student was reading between 20 and 60 correct wpm) for probes was determined. Each student was administered 3 CDP and 3 TORF probes. If the student's median reading performance was not between 20 and 60 correct wpm, the student was administered probes in successively higher or lower levels until the criterion was met. All 13 students in the current study were assessed to be reading at a second grade instructional level; thus CDP and TORF probes used for the current study were derived from passages with second grade 1 month grade level readability to second grade

10 months grade level readability. All CDP were derived from the Scott Foresman (2000) reading series from 2-second grade texts, My Time to Shine and New Beginnings.

Following the determination of instructional level for all students, a different set of three CDP and three TORF probes were administered and scored twice each week for every participant. The administration of probes was counterbalanced. Identical standardized CBM administration and scoring procedures (Shinn, 1989) were used for both CDP and TORF reading probes. Administration was defined as presenting a non-numbered copy of a given reading probe to a student. Scoring was defined as counting the total number of wpm read and subtracting the number of errors from that total. Following the administration of probes, median scores for correct wpm on each probe type (CDP and TORF) were determined. All probes were scored and checked by another scorer for accuracy in scoring. Interscorer agreement was 100%. Interscorer reliability was determined by the following percent agreement formula:

$$\frac{\text{Agreements}}{\text{Agreements} + \text{Disagreements}} \times 100$$

Students were required to read each probe aloud for one minute (Shinn, 1989). The following instructions, taken from Shinn (1989), were given at the beginning of each probe administration: "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 don't know, I'll tell it to you. Be sure to do your best reading. Are there any questions?" Then the student was told to "Start".

As the student read, the examiner followed along in the numbered copy, marking words that were read incorrectly. Words were counted as correct if they were (a) pronounced correctly or (b) self-corrected within three seconds if initially pronounced incorrectly. Errors were defined as mispronunciations, omissions, or substitutions. As Shinn (1989) recommended, repetitions and additions of words were not scored as errors. Finally, if a student hesitated or struggled to pronounce a word for three seconds, the examiner supplied the word and an error was recorded. Following each one-minute reading, a vertical line was placed on the numbered copy after the last word read and the student was thanked.

Data Analysis

Previous work by Powell-Smith and Bradley-Klug (2001) compared reading probe types with the application of individual t-tests. In analyzing their data, Powell-Smith and Bradley-Klug combined all of the generic probe scores to yield a single mean, combined all of the curricular probe scores to yield a single mean, and then compared these two means. This increased the probability of Type II error. Similarly, conducting a series of individual t-tests may have diluted the power of the analysis. A more powerful approach would have been to use a multiple regression procedure, which would have compared all of the individual scores simultaneously while also controlling for time and individual student effects. Pedhazur (1997) suggests, "Multiple Regression analysis (MR) is eminently suited for analyzing collective and separate effects of two or more independent variables on a dependent variable."

The current study employed a Multiple Regression analysis (MR) with a time variable to measure progress during the study period, a dummy variable to detect the effect of the probe type, an interaction term to judge difference in progress by probe type over time, and student dummies to detect individual student effects. A MR was applied due to the power afforded by this procedure (Pedhazur, 1997).

Results

Probes Types Ability to Measure Current Performance

A MR was conducted to examine the differences for each probe type (CDP and TORF) for both current level of performance and progress over time. Descriptive statistics for daily mean wpm scores on CDP and TORF probes over the five-week data collection period are reported in Table 1.

Table 1

Descriptive Statistics For Probe Scores

		Mean wpm		Standard deviation	
Week	Day	CDP	TORF	CDP	TORF
1	Tuesday	38.38	41.38	18.61	20.38
	Thursday	40.00	42.69	18.69	21.19
2	Tuesday	41.76	44.30	19.35	20.49
	Thursday	43.46	44.69	18.36	19.58
3	Tuesday	45.46	46.76	18.86	19.33
	Thursday	46.15	49.38	18.58	20.36
4	Tuesday	48.15	54.15	23.45	23.74
	Thursday	48.69	57.30	19.57	25.38
5	Tuesday	56.30	59.23	22.13	26.16
	Thursday	53.69	58.76	20.84	26.64

Note. wpm = words per minute; CDP = Curriculum Dependent Probes; TORF = Test of Reading Fluency

In interpreting Table 1, TORF wpm means appear to be generally higher than CDP wpm means, however, results of the MR (see Table 2) indicate the difference between probe scores was not statistically significant. Table 1 also indicates that, as a group,

students increased the number of wpm read on both CDP and TORF probes over the course of the five week data collection period; however, the final day of data collection wpm dropped slightly. This minimal decrease in wpm may have been due to the fact that the final data collection took place the day before student's spring break. In interpreting the coefficients, the coefficient on time indicated a gain of .645 wpm per day (see Table 2) and this was significantly different from zero. In examining the results regarding ability to measure current level of performance, no relationship between variables was indicated, thus resulting in failure to reject the null hypothesis. There was no statistically significant difference (at the .05 level) between CDP and TORF probes' ability to measure current level of performance.

Probe Types Ability to Measure Progress Over Time

In examining the results of whether there was a statistically significant difference between CDP and TORF probes in ability to measure progress over time, no relationship was shown, again resulting in failure to reject the null hypothesis. Results indicated no statistically significant difference (at the .05 level) between CDP and TORF probes in ability to measure progress over time. Multiple regression results are reported in Table 2.

Table 2

Multiple Regression with Time and Dummy Variables for Current Level of Performance and Progress Over Time on CDP and TORF Reading Probes

	Coefficients	t Statistic
Intercept	49.601	30.062
time	0.645	12.054
C=1	-1.830	-1.282
time*C=1	-0.114	-1.511
S1	-39.1	- 20.349
S2	30.65	15.951
S3	6.45	3.356
S4	-3.35	-1.743
S5	-27.75	-14.442
S6	-21.95	-11.423
S7	-11.35	-5.907
S8	-25.85	-13.453
S9	-17.3	-9.003
S10	4.5	2.342
S11	10.4	5.412
S12	-36.15	-18.814

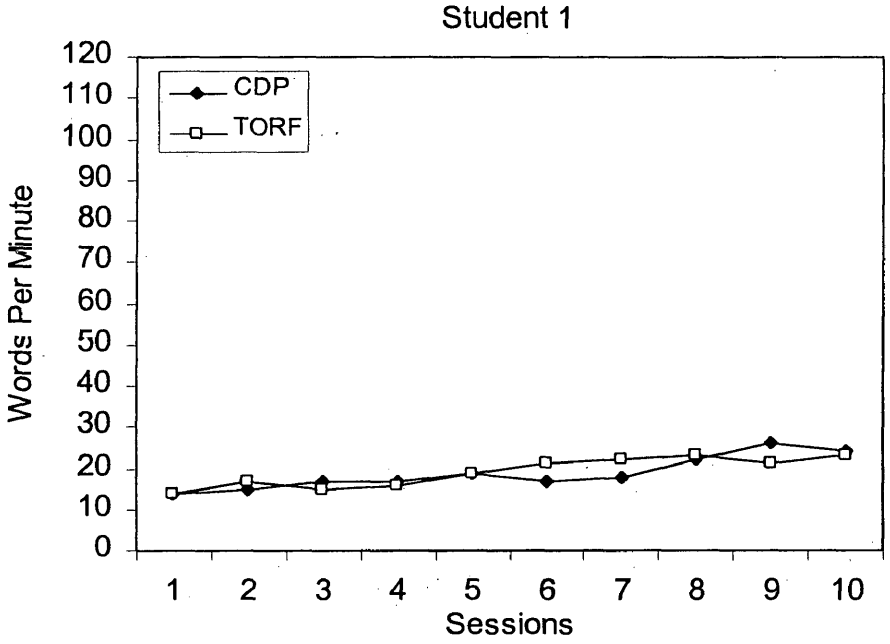
Note. R Square = 0.924

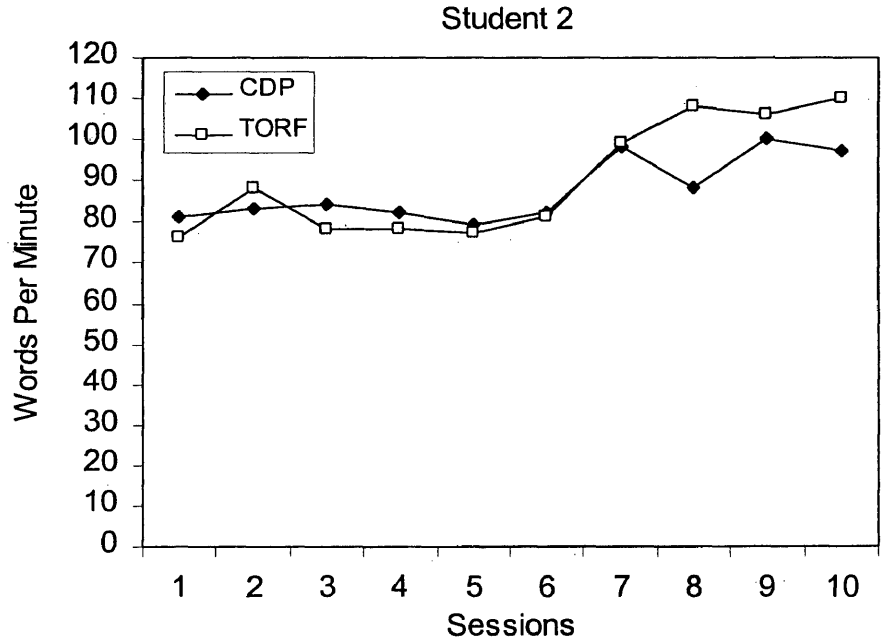
In interpreting Table 2, the dependent variable was performance on the CDP and TORF probes in wpm. The time variable indicated the progress over the study period.

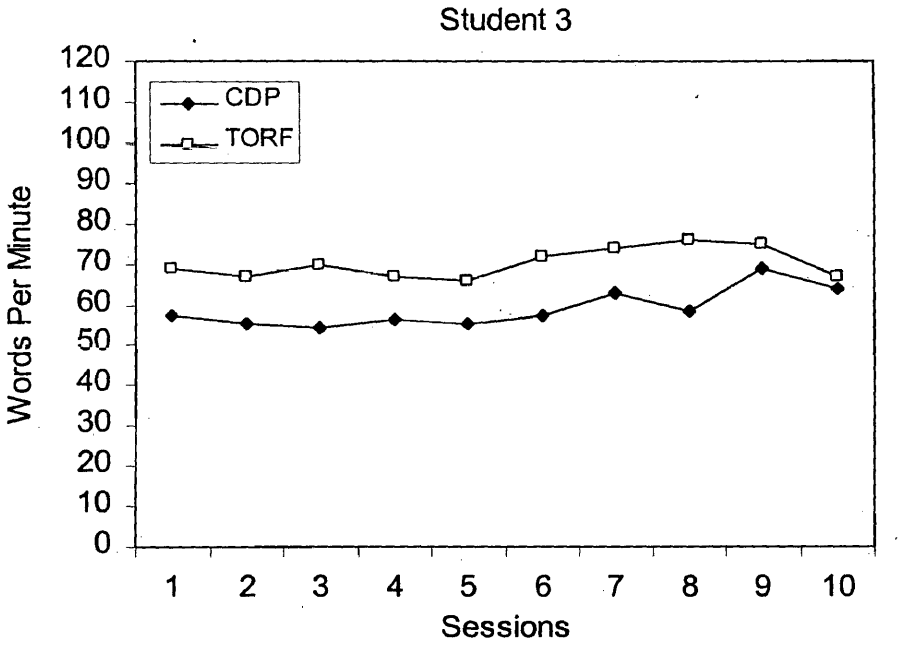
C=1 was the Dummy variable to pick up test effect. C=1*time was the interaction term to indicate whether measured progress differed by probe type (CDP and TORF). S1 through S12 indicated individual student performance and were the student specific dummies.

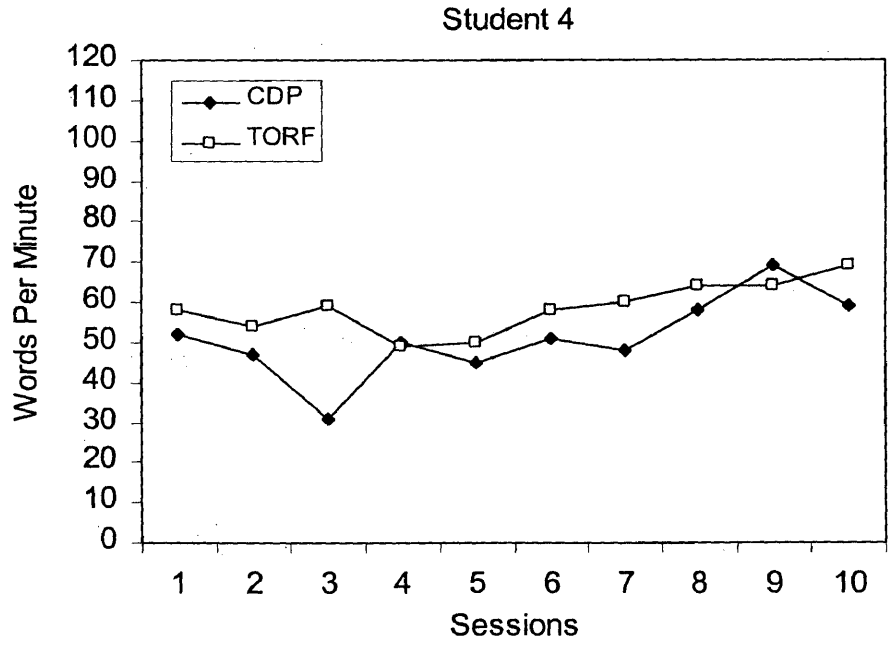
Table 2 indicates, while all students made progress, there was variance in the individual student data (e.g. some students made more progress over the study than others). This variance lends support to the failure to reject the null hypotheses in that regardless of the progress made (many wpm or few wpm) over the data collection period, there was no statistically significant difference between CDP and TORF probes in ability to measure current levels of performance and progress over time.

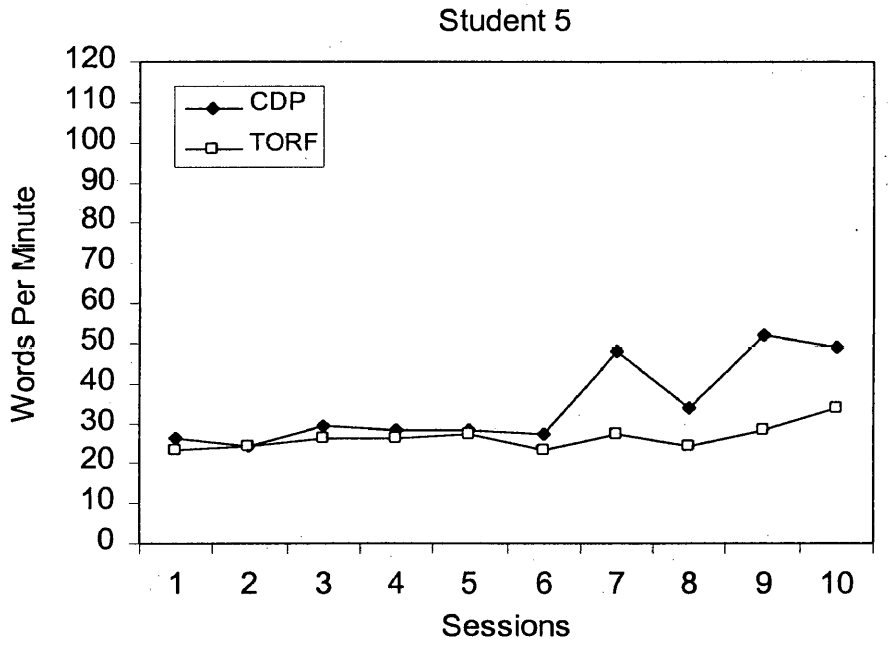
Figures 1 through 13 present individual student data. All figures indicate that every student made progress over the study period as monitored by both CDP and TORF probes. The Figures also indicate that some students read more wpm in CDP probes, other students read more wpm in TORF probes, and still other students made the same amount of progress as measured in wpm in both probe types. While there was some variance in wpm according to probe types the MR in Table 2 indicates that this variance was not statistically significant. In addition, while there was variance in individual data, the MR accounted for individual student differences with the individual student dummy variables. Results of individual student data reported in Figures 1 through 13 lend further support to the failure to reject the null hypotheses in that regardless of the progress made both probe types measured current levels of performance and progress over time similarly.

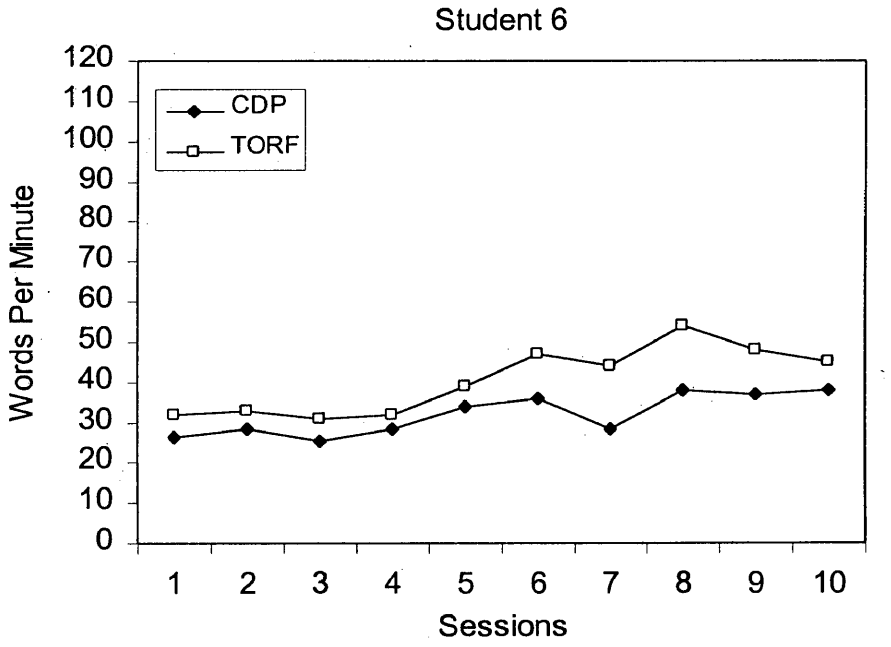

Student 1
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions

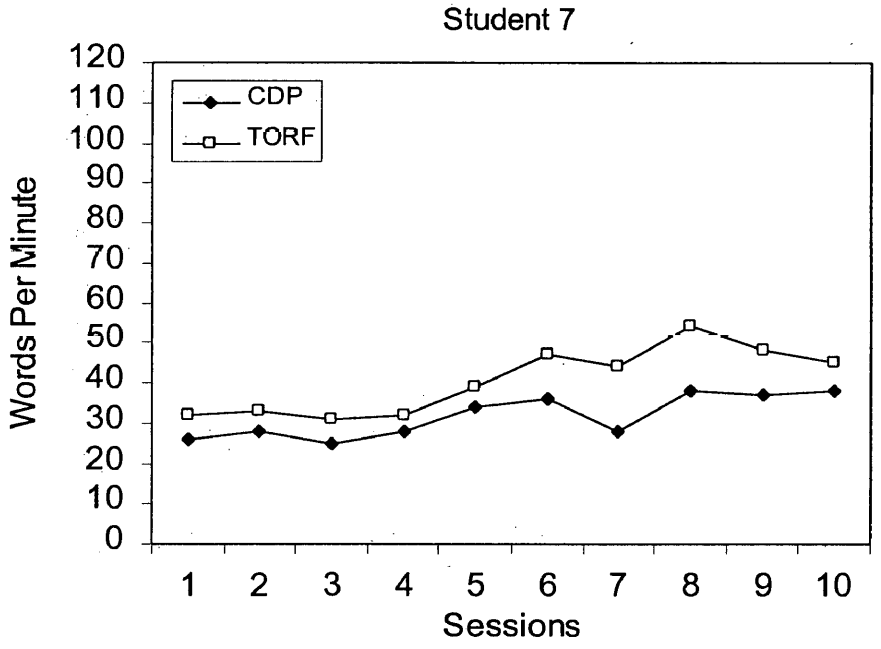

Student 2
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions

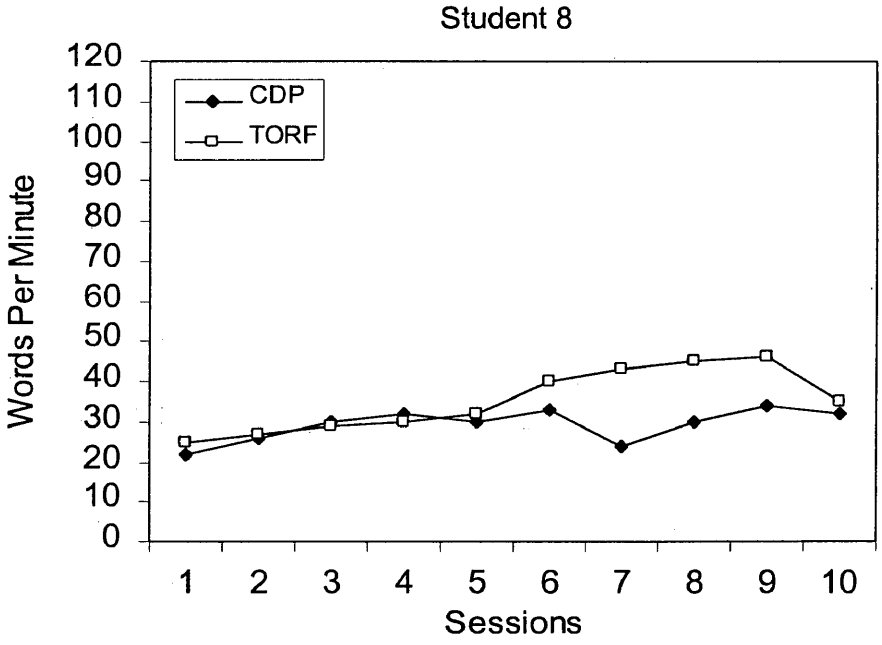

Student 3
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions

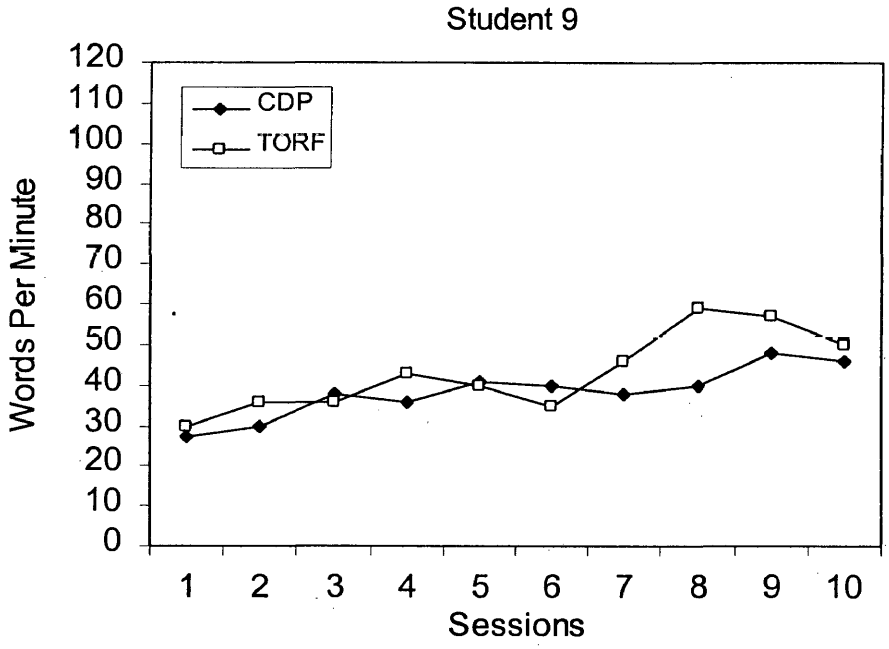

Student 4
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions

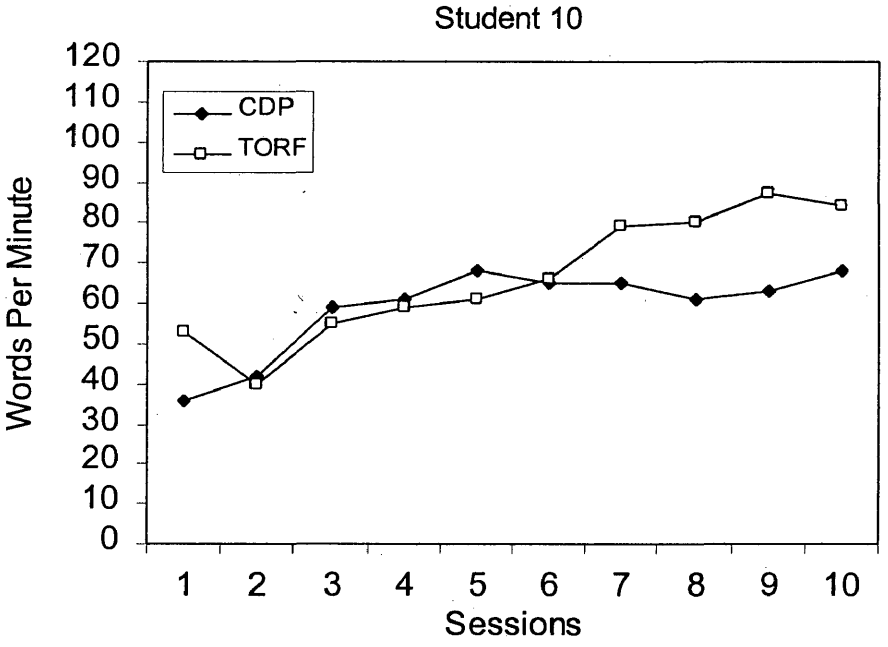

Student 5
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions

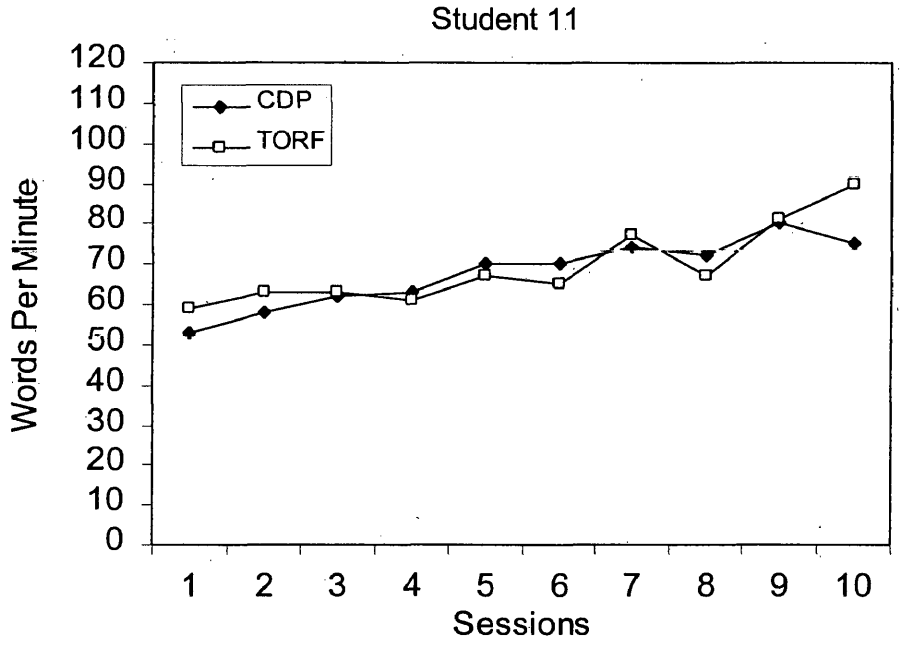

Student 6
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions

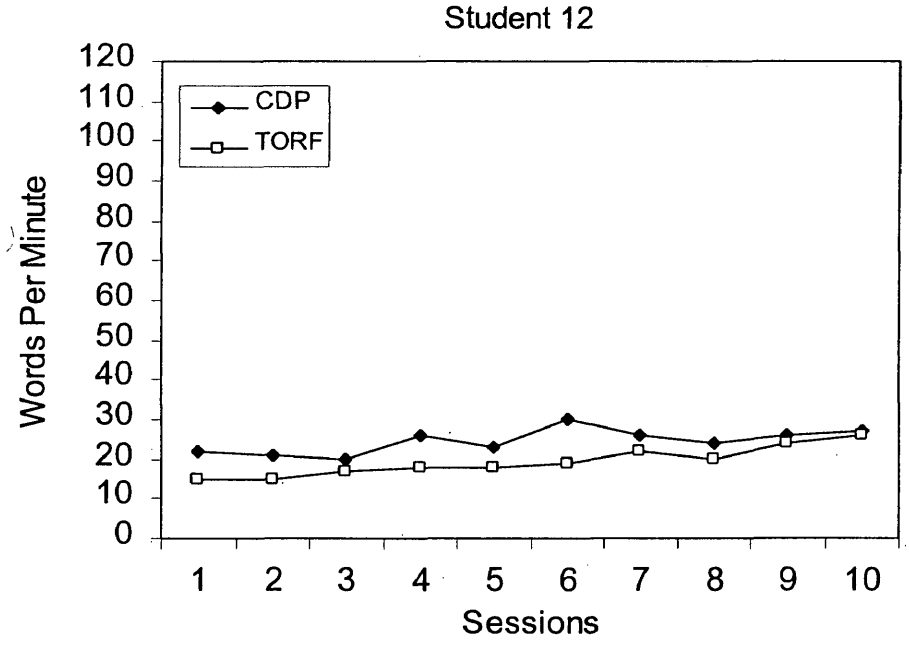

Student 7
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions

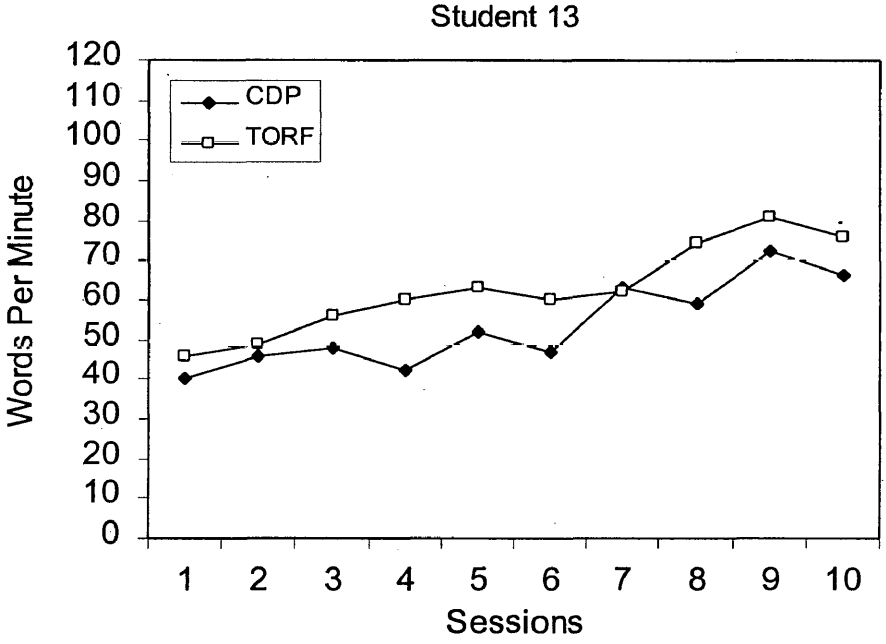

Student 8
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions


Student 9
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions


Student 10
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions


Student 11
120
110
100
90
80
70
60
50
40
30
20
10
0
Words Per Minute
CDP
TORF
1
2
3
4
5
6
7
8
9
10
Sessions


Student 12
120
110
100
90
80
70
60
50
40
30
20
10
0
Words Per Minute
CDP
TORF
1
2
3
4
5
6
7
8
9
10
Sessions


Student 13
CDP
TORF
Words Per Minute
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Sessions

Discussion

The purpose of the present study was to extend the previous research of Powell-Smith and Bradley-Klug (2001) on attempting to determine whether or not CDP and TORF probes vary significantly in capacity to measure current levels of reading performance and whether probe types differ significantly in ability to measure progress over time or slope. Results of the current study indicated no statistically significant difference between CDP and TORF probes in ability to measure current levels of performance. Results also showed no statistically significant difference between the probe types ability to measure progress over time.

These results were similar to, but not a strict replication of, the results of Powell-Smith and Bradley-Klug (2001) who reported a statistically significant difference between measured basal curriculum probes wpm and TORF wpm (i.e., on average, students read more wpm in TORF probes), but no statistically significant difference in the two probe types ability to measure progress over time. The difference between wpm scores may have been the result of Powell-Smith and Bradley-Klug's failure to counterbalance presentation of probe types (specifically, basal/curriculum-dependent probes were always administered first, followed by TORF probes). This presentation may have contributed to a practice effect for the reading of TORF probes. The difference might also have been related to the fact that two different methods were used to place students into probe materials. Specifically, in the Powell-Smith and Bradley-Klug study generic screening probes and selection criteria were used for the generic probes whereas

survey level assessment procedures were used for the basal (curriculum) probes. Finally, in their statistical analysis, Powell-Smith and Bradley-Klug did not employ a MR with a time variable to measure progress during the study period or a dummy variable to account for individual student effects. This decreased the power of their analysis and may have resulted in increased probability of a Type II error.

The current study counterbalanced probe presentation, used the same method for placing students into probe materials, and employed a MR with time and dummy variables. These factors may have contributed to the variation in results indicated in the current study. Over the course of the current study some students read more wpm in CDP, other students read more wpm in TORF probes, and still other students read approximately the same number of wpm in both probe types. In addition, some students made more progress than others over the study period. All students made progress however. Despite the variance in progress, there was no statistically significant difference between probe types in ability to measure current level of performance and progress over time. Results of the current study indicated that both CDP and TORF probes were equally effective in ability to monitor current levels of performance and progress over time.

Implications for Practitioners

The current study's results suggest generic, curriculum-independent reading probes, specifically TORF, are an effectual alternative to probes derived directly from the curriculum. Both probe types were equally effective in measuring current levels of reading performance and reading progress over time. Due to the sample population being

comprised of only second grade students there may be implications regarding the generalizability of results to other grade levels. Second grade readers were selected for the current study due to their potential for growth (Snow et al., 1998) over the data collection period. In the current study of second grade students both CDP and TORF probes were sensitive to current levels of performance and progress over time. Figures 1 through 13 indicate individual student increases in wpm over the study period. Table 1 indicates that, as a group, students increased the number of wpm read on both CDP and TORF probes over the course of the five-week data collection period. The coefficient on time (in Table 2) indicates a gain of .645 wpm per day. This daily gain of wpm is statistically significant from zero.

Consistent with the results of Powell-Smith and Bradley-Klug (2001) all students in the current study made progress over the study period, however, the amount of progress made varied between individual students. This variability served to support the two hypotheses in that regardless of progress made both probe types were equally effective in measuring current levels of performance and progress over time. While the current study employed a MR with a dummy variable to account for individual student differences, the study did not examine why such variability existed between individuals.

Both probe types in the current study included the three critical elements of instructionally valid measurement outlined by Fuchs and Deno (1994). CDP and TORF probes (a) enabled repeated testing over time on material of comparable difficulty, (b)

possessed the capacity to accurately indicate and assess outcomes of instruction, and (c) provided quantitative and qualitative feedback regarding student performance.

Results indicate that teachers, school psychologists and other practitioners would make the same determinations regarding student progress, performance, interventions etc., regardless of the type of reading probes used. Consistent with conclusions by Powell-Smith and Bradley-Klug (2001), results of the current study also suggest a high level of decision reliability across CBM reading probe types.

Several advantages may be associated with the use of generic TORF probes versus CDP probes in the school setting. First, TORF probes provide consistency of measurement materials while controlling difficulty level and maintaining a link to educational outcomes. Second, the entire CBM process may be less labor intensive and considerably faster with the application of TORF probes. This may result in less time and energy commitments for para-educators, teachers and school psychologists if they are not responsible for developing numerous reading probes directly from student's reading curriculum. Third, varied reading curricula could be implemented at all levels to effectively educate based on individual needs without interfering with the application of a single common accountability system for all students. Fourth, concerns regarding promotion of a specific reading curriculum at the classroom, building, and/or district level(s) may be lessened. Finally, the application of generic probes in an entire class, building or school district facilitates the development and subsequent application of local norms.

Limitations and Potential for Further Research

Despite the consistent findings of the current study and the methodological improvements upon Powell-Smith and Bradley-Klug's (2001) investigation, certain limitations of the present study must be acknowledged. Specifically, the setting of a single school in the rural Midwest restricts the range of student characteristics and demographics and may affect the generalizability of outcomes to other schools and other locations.

Another limitation of the current study is that while findings were analyzed as group data, specific individual differences may have been overlooked. An attempt was made to display individual student progress with graphs of each student's data. This individual data were consistent with the direction of the statistical analysis. While the current study employed a MR with a dummy variable to account for individual student differences, the study did not examine why such variability existed between individuals. However, there are some potential hypotheses that may aid in accounting for the variability. Hypotheses include (but are not limited to) that variability may have been due to one or more of the following: type of instruction (e.g. students may receive varied instruction according to classrooms/ teachers), amount of instruction (e.g. some classrooms may have 1 hour of reading instruction daily while others have 30 minutes or some students receive additional instruction at home while others do not), gender (e.g. female students may make more progress than males or vice versa), and primary language or language spoken in the home (e.g. students with a primary language other

than English may make less progress than students with English as their primary language). Variability may also be due to other factors not listed.

With regard to the limitations of the current study, future research may effectively involve a sample population with all or any one of these characteristics: a larger number of students, with readers from diverse backgrounds and locals, and students who are in different grades. Future research may also develop curriculum probes from reading series different from the one included in the current study. In addition, future research may also examine readability programs and devise a more appropriate strategy for placing students in appropriate instructional level probes. Finally, future research may examine individual student differences such as type and amount of instruction, gender, and primary language or language spoken in the home.

Conclusions

Results of this study indicate students' demonstrated short-term progress and growth over time regardless of the administration of CDP or TORF reading probes. The results suggest both types of CBM probes are effective in monitoring students' current levels of performance and progress over time. Similar results were found in Powell-Smith and Bradley-Klug (2001). The results also lend support to and extend previous research that suggests it is not essential that CBM reading probes be derived directly from the curriculum (e.g. Fuchs & Deno, 1994; Powell-Smith and Bradley-Klug, 2001).

Finally, this study provides additional information for para-educators, teachers, school psychologists, and other individuals responsible for monitoring reading progress

in the school setting. When these individuals are determining what type of materials to use for CBM probes, results suggest CDP and TORF probes are comparable in ability to measure current levels of performance and progress over time. Due to the benefits associated with generic probes (in particular TORF probes) such as consistency of measurement materials, decreased time and energy commitment, and the potential for the development of local norms, these probes may provide an effective alternative to probes derived directly from the reading curriculum.

References

Atkinson, R. C., & Jackson, G. B. (1992). *Research and education reform roles for the office of educational research and improvement.* Washington, D. C.: National Academy Press.

Clymer, T. & Fenn, T. (1979). *Ginn Reading 720, Rainbow Edition.* Lexington, MA: Ginn.

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

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

Deno, S.L., Deno, D., Marston, D., & Marston, D. (1987). *Test of reading fluency: Measures for screening and progress monitoring.* Minneapolis, MN: Children's Educational Services.

Deno, S.L., Mirkin, P., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36-45.

Elliot, S. N., & Fuchs, L. S. (1997). The utility of curriculum-based measurement and performance assessment as alternatives to traditional intelligence and achievement tests. *School Psychology Review, 26,* 224-233.

Foresman, Scott (2000). My time to shine. *Second Grade Reading Series.* New York: Addison-Wesley Educational Publishers Inc.

Foresman, Scott (2000). New Beginnings. *Second Grade Reading Series.* New York:

Addison-Wesley Educational Publishers Inc.

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

Fuchs, L. S., & Deno, S. L. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58,* 232-242.

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

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22,* 27-48.

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

Fuchs, L. S., & Shinn, M. R. (1989). Writing CBM IEP objectives. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 130-152). New York: Guilford Press.

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

Habedank Stewart, L. & Kaminski (2002). Best practices in developing local norms for academic problem solving. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology III,* 737-752.

Hargis, C. H. (1987). *Curriculum based assessment a primer.* Illinois: Charles C. Thomas.

Hintze, J. M., Shapiro, E. S., & Lutz, J. G. (1994). The effects of curriculum on the sensitivity of curriculum-based measurement in reading. *Journal of Special Education, 28,* 188-202.

Howell, K. W., Fox, S. L., Morehead, M. K. (1993). *Curriculm-based evaluation: Teaching and decision making (2nd ed.).* Pacific Grove, CA: Brooks/Cole.

Klare, G. R. (1988). The formative years. In B. L. Zakaluk & S. J. Samuels (Eds.), *Readability: Its past, present, and future* (pp. 17-34). Newark, DE: International Reading Association.

Knutson, N., & Shinn, M. R. (1991). Curriculum-based measurement: Conceptual underpinnings and integration into problem-solving assessment. *Journal of School Psychology, 29,* 371-393.

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

Marston, D. (1989). Curriculum-based measurement. What is it and why do it? In M. R. Shinn (Ed.) *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford Press.

Marston, D. (1996). Curriculum-based measurement: An alternative approach to assessment. *Curriculum-based measurement mini-series.* Minneapolis, MN.

Naslund, R. A., Thorpe, L. P., & Lefever, D. W. (1978). *SRA Achievement Series.*

Chicago: Science Research Associates.

Pedhazur, E. J. (1997). *Multiple Regression in Behavioral Research: Explanation and Prediction.* Third Edition. New York: Harcourt Brace College Publishers.

Powell-Smith, K. A., & Bradley-Klug, K. L. (2001). Another look at the "C" in CBM: Does it really matter if curriculum-based measurement reading probes are curriculum based? *Psychology In the Schools, 38(4),* 299-312.

Shapiro, E. S. (1996). *Academic skills problems direct assessment and intervention.* Second Edition. New York: Guilford Press.

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

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

Snow, C., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Standal, T. C. (1987). Computer-measured readability. *Computers in the schools, 4,* 123-132.

Wheeler, G., & Sherman, T. F. (1983, April). Readability formulas revisited. *Science and Children,* 38-40.