



Curriculum-based Measurement of Reading: a critical review

ALISON MADELAINE & KEVIN WHELDALL

Macquarie University Special Education Centre, Macquarie University, Sydney NSW 2109, Australia

ABSTRACT *Curriculum-based measurement (CBM) represents an alternative to both standardised and teacher made reading tests. Oral Reading Fluency (ORF), a curriculum-based measure of reading, is presented as an accurate indicator of both general reading ability and reading comprehension, and as a means of monitoring reading progress towards functional literacy. The relevant literature relating to curriculum-based measurement of reading is critically reviewed, with particular reference to research supporting the use of Passage Reading Tests (PRTs) to measure ORF.*

There is a recognised need for an alternative to traditional measurement tools in education, as many authors have made clear in recent years (see e.g., Fuchs & Deno, 1994; Marston, 1989; Marston, Fuchs, & Deno, 1986; Shinn, 1989). Curriculum-based assessment presents as just such an alternative. This paper will clarify the role of curriculum-based assessment and curriculum-based measurement within the context of reading assessment by examining the available empirical research, discussing relevant current theory and practice, and identifying possible directions for future research.

The Use of Standardised Tests

Marston (1989) has raised the important question of why, given that so many assessment tools are available, do we need an alternative to traditional standardised tests? The answer is that standardised tests are often an inadequate means of accurately indexing academic progress or for making instructional and placement decisions (Fuchs, Fuchs, & Maxwell, 1988; Marston, 1989; Marston, Fuchs, & Deno, 1986; Tucker, 1985). Traditional tests are said to be problematic for a variety of reasons (Fuchs, Fuchs, & Hamlett, 1989b; Galagan, 1985; Jenkins & Jewell, 1993; Marston & Magnusson, 1988; Rodden-Nord & Shinn, 1991; Shinn, 1988).

First, concerns about the technical adequacy of these tests have been expressed (Galagan, 1985; Marston, 1989). Standardised tests are often criticised for a lack of specific content validity, that is, the extent to which they measure the curriculum being taught (Fuchs, Fuchs, & Maxwell, 1988; Fuchs, Fuchs, & Hamlett, 1989a;

Marston, 1989; Marston, Fuchs, & Deno, 1986; Marston & Magnusson, 1988; Rodden-Nord & Shinn, 1991). Content validity is an important consideration when measuring achievement and if this is low, the test results fail to provide an accurate picture of what students have learned (Borg & Gall, 1989).

Second, as standardised tests are typically insensitive to small changes in students' progress, they may indicate no gain when actual improvements have been made (Marston, 1989; Marston & Magnusson, 1988). Marston, Fuchs, and Deno (1986) compared the sensitivity of standardised achievement tests with that of curriculum-based measures. Their preliminary findings indicated that curriculum-related measures were more sensitive to short term gains than were standardised measures in students with learning difficulties.

Third, traditional standardised tests are generally used in pre-post test situations (Marston, Fuchs, & Deno, 1986; Marston & Magnusson, 1988; McCurdy & Shapiro, 1992). This is an inadequate and unreliable means of tracking progress or for making instructional decisions due to the infrequency of assessment and the fact that this represents a summative rather than a formative measure (Marston, 1989). Such forms of assessment only identify ineffective programs after they have been completed. Related to this is the fact that standardised tests quite simply do not typically provide sufficient parallel forms to monitor progress on an ongoing basis (Fuchs, Fuchs, & Hamlett, 1989b; Fuchs, Fuchs, & Maxwell, 1988).

Fourth, standardised achievement tests are often used for a purpose other than that for which they were intended (Galagan, 1985). The purpose of norm referenced or standardised achievement tests is to compare a student against a group of students of the same age or class. This information is then used to make placement decisions (Howell, Fox, & Morehead, 1993). Problems arise when standardised tests are used to measure learning when they were designed to measure relative standing in a group (Marston, Fuchs, & Deno, 1986). Such test results are typically expressed in age or grade equivalent form and have very limited utility for making instructional decisions (Marston, 1989; Rodden-Nord & Shinn, 1991; Shinn & Hubbard, 1992).

Fifth, standardised tests are perceived to be inadequate for making instructional decisions, *particularly by special educators*. Thurlow and Ysseldyke (1982) conducted a study in which special educators and school psychologists were asked to rate a set of standardised measures in terms of instructional usefulness. The results indicated variations in the perceptions of the two groups, with school psychologists seeing such tests as more instructionally useful. The authors concluded that special educators may need alternative assessment procedures.

These difficulties associated with standardised measures present particular problems for professionals in the face of increasing demands for accountability and even legal requirements (Marston, Fuchs, & Deno, 1986; Marston & Magnusson, 1988; Reschly, Kicklighter, & McKee, 1988a, 1988b). Marston (1989) cites cases of education-related litigation in the United States in which the need for alternative methods of assessment was identified as a result of problems with current assessment procedures.

Other disadvantages to the use of standardised measures include the fact that they are typically too time consuming to administer routinely (Fuchs, Fuchs, & Hamlett,

1989b), there is an emphasis on selection-type items (with students being required to match or point rather than produce), they are often expensive to administer (Marston & Magnusson, 1988), and they usually do not measure fluency, an important consideration, particularly with respect to reading (Carver, 1992; Fuchs, Fuchs, & Maxwell, 1988; Marston, 1989; Shinn, Knutson, Good, Tilly, & Collins, 1992).

The Assessment of Reading

Reading is typically measured either by using published, standardised reading tests or informal teacher-made tests or observations. Several authors have highlighted the difficulties associated with using standardised reading tests to assess reading proficiency for the reasons outlined above about tests in general (Fuchs, Fuchs, & Hamlett, 1989b; Jenkins & Jewell, 1993), even though their technical, psychometric properties may be well established. The technical characteristics of teacher-made tests, however, are unknown, rendering their usefulness as assessment tools questionable at best (Fuchs & Deno, 1991; Jenkins & Jewell, 1993; Marston & Magnusson, 1988), even though, superficially, they may present as having greater face validity. Jenkins and Jewell (1993) report instances where teacher judgment of students' reading ability correlated positively with their performance on standardised tests, but there was considerable variability in the accuracy of these judgments, with correlations ranging from almost zero (i.e., no correlation) to .89. Curriculum-based measurement of reading is said to be a better indicator of general reading ability and progress (Fuchs, Deno, & Mirkin, 1984; Potter & Wamre, 1990). What follows is an explanation of curriculum-based measurement in general and, more specifically, how it relates to assessment in reading.

Curriculum-based Measurement

Curriculum-based measurement (CBM) has been described as a methodology for indexing student proficiency in the curriculum (Fuchs & Deno, 1992). It has been used for a variety of purposes including progress monitoring, screening, referral, and instructional decision-making (Deno, 1987; Deno & Fuchs, 1987; Fuchs & Deno, 1992; Mehrens & Clarizio, 1993). CBM is becoming increasingly more popular as a tool for indexing reading progress in particular (Deno, 1985), and has been used in a variety of academic and non-academic areas such as spelling, mathematics, social adjustment, and written expression (Deno, 1985; Deno, Marston, & Mirkin, 1982; Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, Hamlett, & Allinder, 1991; Marston & Magnusson, 1988). This paper, however, will limit discussions of CBM mainly to the area of reading.

Although definitions of CBM are clear, confusions still exist about exactly what CBM is referring to and, in particular, how it differs from curriculum-based assessment (CBA) (Fuchs & Deno, 1991). Put simply, CBM focuses on long-term instructional goals whereas CBA focuses on a series of short-term instructional objectives. In order to define CBM adequately, its relationship to CBA needs to be

discussed. CBA has been defined by Tucker (1985) as an assessment procedure that draws test stimuli from the students' curricula with repeated testings across time and where the assessment information thus provided is used in instructional decision-making (Fuchs & Deno, 1991). CBM is a form of CBA, but there are distinct differences between the two. These distinctions can be more clearly seen by looking at how CBA and CBM fit into two measurement models: specific subskill mastery measurement (SSMM) and general outcome measurement (GOM) (Fuchs & Deno, 1991). Most forms of CBA, including CBM, can be classified as either SSMM or GOM (Fuchs & Deno).

CBA usually relies on SSMM. Within this type of measurement, the curriculum is presented as a series of criterion referenced items in the form of ordered instructional objectives (Fuchs & Deno, 1991). This form of assessment is useful if the intention is to focus on mastery of individual skills. SSMM can be problematic, however, in a number of ways. First, measurement is directed by the instructional hierarchy, which may or may not reflect an appropriate sequence of skills. Second, each time a new skill is mastered, there may need to be a measurement shift, which could influence the results due to the possibility of students having to adjust to a new task and a different response format. Third, SSMM often requires that teacher made tests are used, which possess unknown reliability, validity and other technical characteristics (Fuchs & Deno, 1991).

Alternatively, GOM assesses proficiency on long term goals. Instead of relying on a hierarchy of skills to determine measurement, a specific domain is identified (in the case of reading, this is usually oral reading fluency) and is used to measure proficiency. Because GOM assesses progress towards broader goals, it is sensitive to instructional effects. In contrast to the measurement shifts that may be necessary with SSMM, the constant level of difficulty and task consistency of GOM ensure that such shifts are not needed (Fuchs & Deno, 1991). Moreover, and perhaps this is one of the most obvious differences between SSMM and GOM, GOM relies on a prescriptive measurement methodology that has been standardised (Fuchs & Deno, 1991, 1992; Mehrens & Clarizio, 1993). This removes many of the unknown effects of reliance on teacher made tests.

CBM is one example of GOM. It is a blend of traditional and curriculum-based assessment models in that it is based in the curriculum, but at the same time it is standardised in much the same way as a norm referenced test would be. As Knutson and Shinn (1991) note, CBM can be curriculum, norm, and individually referenced all at the same time. CBM originally developed from earlier work on a measurement system known as Data-Based Program Modification (DBPM) (Deno, 1987). This assessment model originally used SSMM, but the inadequacies of this approach, as outlined above, created the need for a measurement system that was reliable and valid, quick and easy to administer repeatedly, inexpensive, unobtrusive, sensitive to small changes in progress, and able to be used to make instructional decisions (Deno, 1985; Deno, Mirkin, & Chiang, 1982; Fuchs & Deno, 1991). A study investigating the efficacy of DBPM by Fuchs, Deno, and Mirkin (1984), found that students achieved better results when their teachers employed such ongoing measurement and evaluation techniques. This finding was subsequently supported

by the results of a meta-analysis of the effects of systematic formative evaluation (Fuchs & Fuchs, 1986).

Empirical research has consistently demonstrated the effectiveness of curriculum-based measurement in instructional decision-making (Deno, Mirkin, & Wesson, 1984; Fuchs, Fuchs, & Hamlett, 1989a; Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994). An appropriate long range goal and performance criterion for a skill are determined using results from the initial assessment (Deno, Mirkin, & Wesson, 1984). While the task remains the same during frequent assessment, it is the performance criterion that represents achievement of the goal. This system makes weekly or twice weekly monitoring, for example, very easy. By plotting data on a graph, it can be clearly seen if the student is making progress towards the long range goal or if an instructional change needs to be made. Moreover, simple linear graphing is employed obviating the necessity for the complexities of semi-log graphing typically found in precision teaching, for example.

There are several advantages to the use of CBM. These include: increased content validity because CBM tests what is taught (Deno & Fuchs, 1987); indexed growth in mainstream areas—which is particularly useful in special education where the goal is for students to function within a mainstream curriculum (Deno & Fuchs); clearly communicable measurement data (Deno & Fuchs); a production-type response format (Marston & Magnusson, 1988); frequent measurement and therefore more effective monitoring (Marston & Magnusson, 1988); sensitivity to small changes in progress; multiple, parallel forms; shorter and therefore cheaper administration times; and a measure of fluency rather than accuracy alone. These advantages address many of the disadvantages of standardised tests as outlined above.

Curriculum-based Measurement and Reading Assessment

As stated above, this paper will look mainly at the area of reading. Much of the research on CBM concerning reading has focused on determining the best measure of reading progress (Deno, 1985; Deno, Mirkin, & Chiang, 1982). The available research indicates that oral reading fluency (ORF) is an accurate measure of general reading ability including reading comprehension (Carver, 1992; Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Jenkins & Jewell, 1993; Parker, Hasbrouk, & Tindal, 1992; Shinn *et al.*, 1992). Oral reading fluency represents a combination of accuracy and rate. It is typically measured by a Passage Reading Test (PRT) and is usually expressed as words read correctly per minute (Hasbrouk & Tindal, 1992). In a series of studies conducted by Deno, Mirkin, and Chiang in the early 1980s, several possible measures of reading progress (e.g., reading words in isolation, oral reading, cloze comprehension, and word meaning tasks) were identified and compared with the results of student performance on various standardised reading tests (Deno, Mirkin, & Chiang, 1982). The results indicated high correlations between reading aloud measures and standardised reading tests, correlations varying between 0.73 and 0.91, with most values over 0.8. This led to the conclusion that reading aloud measures, as exemplified by a Passage Reading Test

(PRT), may be a valid means of measuring competence and indexing progress in reading.

A PRT typically requires students to read from a passage at grade level for 1 min. The examiner marks errors on a numbered copy as the student reads from the unnumbered copy. If the student does not know a word, the examiner supplies the word after 3 s and scores one error. For more details on administration and scoring procedures, see Shinn (1989). After the student has completed reading the passage for 1 min, words read correctly per minute is simply calculated by subtracting the number of errors made from the total number of words read, to give words read correctly per minute.

Research on Oral Reading Fluency

The score obtained from a PRT, that is, the number of words read correctly in 1 min, is a measure of oral reading fluency (ORF). ORF has been used synonymously with “rapid decoding” (Shinn *et al.*, 1992). Other studies found high correlations between reading aloud measures and reading comprehension (both literal and inferential) measures on standardised reading tests. Deno, Mirkin, and Chiang (1982) found higher correlations between oral reading tasks and standardised comprehension tests (0.78 and 0.8), than between cloze comprehension measures and standardised tests of comprehension (0.67 and 0.71). Similarly, Fuchs, Fuchs, and Maxwell (1988) found stronger correlations between oral reading scores and standardised tests of reading comprehension than three informal measures of comprehension (question answering, written recall, and written cloze) and the standardised tests. These results suggest that there is more to oral reading fluency than just rapid decoding. It appears to be a good performance indicator of comprehension ability as well as general reading ability (Fuchs, Fuchs, & Maxwell). Jenkins and Jewell (1993) found similar results. These studies provide evidence of the criterion validity of ORF.

Despite research evidence suggesting that oral reading fluency gives an accurate indication of comprehension in addition to general reading proficiency (Fuchs, Fuchs, & Maxwell, 1988), it is recognised that face validity of this performance indicator is low (Deno, 1985; Fuchs, Fuchs, & Maxwell, 1988; Parker *et al.*, 1992; Potter & Wamre, 1990; Shinn *et al.*, 1992). Potter and Wamre suggest that face validity can be increased if educators witness the successful use of PRTs to measure general reading ability and comprehension in the classroom. It appears that this may only be achieved by teachers comparing scores on PRTs with their own judgment about students’ reading ability. This more qualitative comparison may be necessary as quantitative evidence of the criterion validity of this measure already exists (e.g., Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Jenkins & Jewell, 1993).

In an attempt to increase the face validity of oral reading fluency, Parker *et al.* (1992) compared the criterion validity of traditional ORF with two new types, both modified to include miscues. The first type counted only meaning change errors and the second type resembled traditional oral reading accuracy. This was expressed as

words read correctly per passage, without any consideration of time taken. Like the modified ORF measure, the accuracy measure only included meaning change errors. Interestingly, there was little difference between traditional oral reading fluency and the modified measures in differentiating between students in regular versus special education. It is difficult to imagine how presenting teachers with many different measures could improve the face validity of oral reading fluency and it is, therefore, not surprising that the authors concluded that traditional oral reading fluency seems more practical for use in the classroom.

A number of studies have also looked at the relationship of CBM to theoretical reading models. Potter and Wamre (1990) examined whether CBM is congruent with developmental models of reading, in particular, Chall's (1983) stages of reading and LaBerge and Samuels' (1974) model of automaticity in information-processing. Decoding speed is seen as an important aspect of both models. According to these models, as a student develops automaticity in decoding, more of their attention can then be devoted to comprehension. This would account for the positive correlation between increasing oral reading fluency and improved comprehension. A study conducted in 1992 by Shinn *et al.* used confirmatory factor analysis to verify that oral reading fluency is indeed compatible with current theoretical models of reading.

Researchers have also attempted to establish norms for oral reading fluency in order to provide educators with an indication of how far above or below "the average" an individual student may be said to be. Marston and Magnusson (1988) established a set of local norms for a group of Public Schools in Minneapolis, as did Shinn (1988) for a different group of schools. Hasbrouk and Tindal (1992) subsequently reported on several inadequate attempts at determining norms for oral reading fluency as well as an attempt of their own. They assessed up to 9,000 students from Years 2 to 5 over a nine year period. Students were required to read grade level passages regardless of instructional level (see p. 42 of Hasbrouk and Tyndal for the norms produced). One limitation of this study, however, was that the curriculum accessed was not constant across schools included in the study. Although it may appear that any differences between reading curricula could be said to be controlled for, over so many schools, a more accurate way of defining norms might be to use a generic set of passages across all participating schools. We will return to this issue later in this paper.

Mehrens and Clarizio (1993) have criticised these types of measures used to indicate progress in a particular academic area. They see fluency, for example, as a limited sample of reading behaviour and yet there is a growing body of research that supports the use of oral reading fluency as an accurate predictor of overall reading proficiency (Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Shinn *et al.*, 1992). They go on to mention the irrelevance of these measures to the secondary curriculum but if students have not reached a certain level of skill in a particular area, for example, functional literacy, by the time they enter secondary school, these measures assume enormous relevance. Similarly, it may be possible to develop other curriculum-based measures that are relevant to the secondary curriculum. Whatever the case, this is not to deny their utility in the primary grades to monitor the performance of, say, low progress readers.

Reliability

As CBM relies on a standardised set of procedures, research in the area of CBM in reading has also focused on developing a set of procedures for administering and scoring passage reading tests (Deno, 1987). It is important that procedures be consistent across testings and data collectors in order to achieve high reliability. (See Shinn, 1989 for a detailed exposition of administration and scoring procedures.)

The use of curriculum-based measurement in reading needs to take into account both alternate forms reliability and interscorer agreement (Deno & Fuchs, 1987). As data collection is relatively easy, with one consistent task there should be high interscorer agreement, and this has been shown to be the case (c.g., Wheldall & Madelaine, 1997). Alternate forms reliability coefficients between parallel passages can be readily calculated by correlating students' scores across parallel passages. Moreover, as curriculum-based measures are administered frequently, the standard error of measurement is decreased and reliability is increased (Marston & Magnusson, 1988).

The Role of Curriculum in CBM

Clearly, and by definition, curriculum is of paramount importance within CBM. Fuchs and Deno (1992) have examined the effects of curriculum differences on the technical features of CBM in reading. Here, curriculum referred to the grade level or level of difficulty of the passages and also to the basal reading series used. Their study looked at both criterion validity and developmental growth rate. The authors argued that the effects of varying curriculum on the criterion validity of oral reading fluency is relatively unknown. The results of their study found, perhaps surprisingly, that criterion validity was not influenced by the difficulty level of text employed for the PRTs (Fuchs & Deno, 1992).

A 1994 study by Hintze, Shapiro, and Lutz also looked at the effects of curriculum on CBM in reading. They compared the sensitivity of passages selected from a basal reading program to those taken from a literature-based reading program. Their results indicated that passages taken from literature-based programs were less sensitive to changes in reading proficiency over time than those drawn from the basal reading series. This is not an unexpected result, since the vocabulary is generally more controlled in basal readers.

A study conducted by Jenkins and Jewell (1993) found significantly high correlations between passage reading and two standardised tests of reading (0.80 and 0.88). The study included Years 2 to 6 and correlations based on grade level were calculated. Results showed a decrease in the correlation coefficients as grade level increased. The authors suggest that some sort of ceiling effect may affect the results at higher grade levels which seems highly probable.

Swain and Allinder (1996) compared the sensitivity of two curriculum-based measures, oral reading and computer maze (an activity similar to cloze, except alternatives are supplied), to evaluate the progress resulting from using the repeated reading technique. Two advantages of using computer maze are that the students in

the study preferred it and that data was collected and feedback given by the computer, thus saving teacher time. The results revealed, however, that when repeated reading was introduced, significant improvement showed up on the oral reading measure but not on the computer maze measure. These results have important implications for educators. First, instructional decisions made on the basis of the two measures could easily have been very different. Second, using computers to assist with data collection in this way, for the sake of convenience, may be dangerous and ultimately counter productive.

When considering what constitutes an appropriate measure of reading progress, the level of difficulty of the measure must be considered, particularly when this measure will be used to monitor progress towards a long-term goal. Mirkin and Deno (cited in Fuchs & Deno, 1992) measured student performance using text at three difficulty levels. They found that material at frustration level was less sensitive to changes in progress than material at instructional or independent levels. This finding has implications for determining the minimum in an optimal reading range associated with a particular measure; for example, at what reading level does the instrument start to measure reading effectively? This is to say that there may be a floor effect associated with some measures for some students. Consequently, if we accept Fuchs and Deno's criterion for determining frustration level text, that is, reading only 10 to 30 words correctly per minute, employing such a passage (on which a student reads fewer than 30 words correctly per minute) would not be as sensitive a tool for measuring small changes in reading proficiency. If students are reading a passage at a rate of fewer than 30 words per minute, the passage may still be appropriate for pre- and post-testing, but not for frequent monitoring as the student may become increasingly distressed at having to read material at frustration level on an ongoing basis.

With regard to the source of the passages used (usually a basal reading series), Fuchs and Deno (1992) refer to Taylor, Willits and Richards (1988), who queried whether curriculum bias may be a cause for concern. (By definition CBA and, therefore, CBM must be curriculum-based.) Fuchs and Deno (1994) also subsequently considered Tucker's (1987) concern that measurement must be based on the materials in which instruction occurs in order to be considered instructionally useful. More recent findings, however, suggest that drawing material from within the student's specific, actual curriculum need not (and perhaps even should not) be a critical feature of CBM. Mehrens and Clarizio (1993) argue that it is the content validity that is important rather than the fact that measures are curriculum-based. Moreover, not only is it the case that using a student's specific curriculum may not be necessary, but also using it may have a number of disadvantages (Fuchs & Deno, 1994).

One disadvantage is the possible differences in levels of difficulty across different basal reading programs (Duffelmeyer, 1983; Fuchs & Deno, 1994). Fuchs and Deno report a study carried out in 1982 by Tindal and colleagues which compared performance levels across reading programs and found that there were differences in readability levels. These differences, however, did not influence the rank ordering of students in the study. The authors concluded that provided a student was tested

across time within the same reading curriculum, differences between curricula would not affect indications of progress. This is a valid assumption only if the reading curriculum remains the same and the student does not change schools and, therefore, curriculum.

A related disadvantage is the possibility of differences within a basal reading series. Fuchs and Deno (1994) recognise that there can be substantial variations in the readability level within the same book, let alone books supposedly at the same level, and that these variations will affect measurement. Moreover, when passages are drawn from a basal reading series, they inevitably represent only a portion of a story. This means that when the student reads the passage, they may start half way through a story. This may affect reading comprehension and therefore accuracy and rate, or fluency, since they appear to be so inter-related. Fuchs and Deno acknowledge that this feature of passages sampled directly from reading curricula could affect comprehension during measurement.

A final disadvantage of strict curriculum-based measurement is the risk associated with using the same material for teaching and testing: the student may be familiar with the passage used to indicate progress. This is dependent on whether the passage being used for testing was recently used for instructional purposes (Fuchs & Deno, 1992).

Uses of Passage Reading Tests

Clearly, passage reading tests have a multiplicity of uses and they have indeed been employed in a number of ways, as the following three examples may illustrate. In a study by Wheldall, Center, and Freeman (1993), passage reading tests formed part of an assessment battery used to investigate the effectiveness of the Reading Recovery Program in Sydney primary schools. As this was an experimental study, the passages were used at pre-test, post-test, and maintenance stages as well as being used to track progress every fortnight. The data collected clearly demonstrated the increasing gains made by the Reading Recovery group over the control group, a trend which was consistent with that shown by data collected with other assessment tools employed in the study.

Rodden-Nord and Shinn (1991) used passage reading tests in a study designed to investigate the academic performance ranges in reading, with identification of students likely to be referred to special education services in mind. As would be expected, there was a wide range of reading ability found, particularly in the higher grades.

In a double blind study by Stoner, Carey, Ikeda, and Shinn (1994), CBM was used to determine the most appropriate dose of methylphenidate (Ritalin) for students diagnosed with attention deficit hyperactivity disorder. The authors looked at whether students' maths and reading scores would vary with changes in the dose of methylphenidate. They found that the change in dose affected reading and maths scores and concluded that the use of CBM in stimulant medication trials is promising.

CBM and Functional Literacy: Recent Research and Developments

The expression “long term” is frequently employed when discussing the reading goals toward which CBM seeks to monitor progress and this can be a little misleading. As previously argued CBM measures the end product *per se* (i.e., reading words in context in continuous prose) rather than specific subskills. Such goals frequently target the end of a school year or year of instruction (Fuchs, 1989; Fuchs & Deno, 1991), but if we are seeking to measure proficiency in the global outcomes toward which the entire curriculum is directed, as stated by Fuchs and Deno, simply measuring to the end of a school year may be inappropriate and inadequate. This is particularly true in reading where it is useful to measure progress over a number of years, and even more pertinent in the case of low progress readers needing to make above average progress in order to catch up. Galbraith and Clayton (1997), for example, investigated the oral reading fluency of Year 4 students using grade level passages taken from the curriculum. This is useful information from the point of view of gaining information about where the students are currently located within their curriculum, as a means of determining resource support, and also for tracking short-term reading improvement. It is, arguably, less effective, however, to direct this level of effort into the collection of data that cannot be used to track the reading progress of such students over a number of years. If reading progress is indexed on grade level passages each year, accurate comparisons cannot be made across years, as Rodden-Nord and Shinn (1991) have emphasised.

Perhaps a better way of measuring progress toward long-term goals in reading is to associate a long-term goal with a minimum level of literacy, or a level of “functional literacy,” rather than the end of a school year, again particularly when considering older low progress readers. There are numerous definitions of functional literacy but, as several authors recognise, these are vague at best (Graff, 1995; Luke & Gilbert, 1993; Radencich, 1994). If we are to use functional literacy to represent an outcome or long-term goal, then it needs to be very clearly defined in terms of a level of reading proficiency, possibly linked to a reading age. There appears to be little research attaching a specific reading level to functional literacy, however, except for a study by Goyen (1977) which found functional literacy to be apparent at a reading age of approximately 10 years 3 months.

Most of the research looking at reliability and validity of PRT measures has used Year 3 passages with students of differing ages (Fuchs & Deno, 1992). If this age is representative of a minimum level of functional literacy, however, then using Year 3 passages as a tool to measure progress towards a long-term goal is inadequate. If we assume, as our best guess and for the sake of argument, that a reading age of around 10 years 3 months approximates functional literacy, students reading at or near this level will find Year 3 passages too easy and the passages will not be sensitive to small changes in progress.

If we assume that curriculum-based measures, in this case passage reading tests, need not be taken directly from a reading curriculum, such tests could comprise a series of specially written, generic passages for monitoring progress towards functional literacy. Such passages would need to satisfy all of the criteria for any good

curriculum-based measure. In addition to being quick and easy to administer and score, they would need to have multiple parallel forms, be reliable, and valid. As mentioned above, the readability levels of passages taken directly from a reading curriculum is of concern due to the high likelihood of considerable variability between passages, as Duffelmeyer (1983) (among others) has maintained. One advantage of using a generic set of passages is that such variations can be deliberately minimised.

Such a set of passages is currently being developed at Macquarie University Special Education Centre (e.g., Wheldall, 1996; Wheldall & Madelaine, 1997). Preliminary investigations employing both regular and low progress readers have concentrated on establishing consistent readability/difficulty level across and within five passages, from an initial selection of 14, written especially for this purpose. Each 200 word passage comprises a complete story. Internal consistency of passages has been estimated by comparing words read correctly in the first minute to words read correctly per minute averaged over the whole passage. Preliminary data indicate very high correlations both within and between the five passages, averaging around 0.95. These passages are a promising assessment tool but will need considerable further research and development.

In this literature review, the problems associated with using standardised reading tests to make instructional decisions and to track reading progress have been considered. The research reviewed suggests that curriculum-based measures of oral reading fluency using passage reading tests may be an effective, alternative way to measure progress towards a long-term reading goal such as functional literacy, particularly with low progress readers. The traditional method of selecting and using such passage reading tests to measure reading progress has been shown to be problematic, however. A set of generic, specially written passages at an appropriate level, such as those described above, might address the problems associated with traditional passage selection and shows promise as a method of tracking reading progress that is quick, simple, reliable, valid, and inexpensive. If this promise were to be realised, special educators and other teachers would be provided with one of the necessary prerequisites for the effective instruction of low progress readers.

References

- BORG, W.R. & GALL, M.D. (1989). *Educational research: An introduction* (5th ed.). New York: Longman.
- CARVER, R.P. (1992). What do standardised tests of reading comprehension measure in terms of efficiency, accuracy, and rate? *Reading Research Quarterly*, 27, 347–359.
- CHALL, J.S. (1983). *Stages of reading development*. New York: McGraw-Hill.
- DENO, S.L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219–232.
- DENO, S.L. (1987). Curriculum-based measurement. *Teaching Exceptional Children*, 20, 41–42.
- DENO, S.L. & FUCHS, L.S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children*, 19(8), 1–16.
- DENO, S.L., MARSTON, D. & MIRKIN, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children*, 46, 368–371.

- DENO, S.L., MIRKIN, P.K. & CHIANG, B. (1982). Identifying valid measures of reading. *Exceptional Children*, 49, 36–45.
- DENO, S.L., MIRKIN, P.K. & WESSON, C. (1984). How to write effective data-based IEPs. *Teaching Exceptional Children*, 16, 99–104.
- DUFFELMEYER, F.A. (1983). Response to Fuchs, Fuchs and Deno. *Reading Research Quarterly*, 18, 370–371.
- FUCHS, L.S. (1989). Evaluating solutions: Monitoring progress and revising intervention plans. In M.R. SHINN (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 158–181). New York: The Guilford Press.
- 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., DENO, S.L. & MIRKIN, P.K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal*, 21, 440–460.
- FUCHS, L.S. & FUCHS, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53, 199–208.
- FUCHS, L.S., FUCHS, D. & HAMLETT, C.L. (1989a). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education*, 10(2), 43–52.
- FUCHS, L.S., FUCHS, D. & HAMLETT, C.L. (1989b). Monitoring reading growth using student recalls: Effects of two teacher feedback systems. *Journal of Educational Research*, 83, 103–110.
- FUCHS, L.S., FUCHS, D., HAMLETT, C.M. & ALLINDER, R.M. (1991). The contribution of skills analysis to curriculum-based measurement in spelling. *Exceptional Children*, 57, 443–452.
- 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, 518–537.
- FUCHS, L.S., FUCHS, D. & MAXWELL, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, 9(2), 20–28.
- GALAGAN, J.E. (1985). Psychoeducational testing: Turn out the lights, the party's over. *Exceptional Children*, 52, 288–299.
- GALBRAITH, E. & CLAYTON, M. (1997). An investigation of the oral reading fluency of grade 4 students. *Australasian Journal of Special Education*, 21, 98–114.
- GOYEN, J.D. (1977). *Adult illiteracy in Sydney*. Canberra: Australian Association of Adult Education.
- GRAFF, H. (1995). *The labyrinths of literacy: Reflections on literacy past and present*. Pittsburgh, PA: University of Pittsburgh.
- HASBROUK, J.E. & TINDAL, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children*, Spring, 41–44.
- HINTZE, J.M., SHAPIRO, E.S. & LUTZ, J.G. (1994). The effects of curriculum on the sensitivity of curriculum-based measurement in reading. *The Journal of Special Education*, 28, 188–202.
- HOWELL, K.W., FOX, S.L. & MOREHEAD, M.K. (1993). *Curriculum-based evaluation: Teaching and decision making* (2nd ed.). Pacific Grove, CA: Brooks/Cole.
- JENKINS, J.R. & JEWELL, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children*, 59, 421–432.
- KNUTSON, N. & SHINN, M. (1991). Curriculum-based measurement: Conceptual underpinnings and integration into problem solving assessment. *Journal of School Psychology*, 29, 371–393.
- LABERGE, D. & SAMUELS, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.

- LUKE, A. & GILBERT, P. (1993). *Literacy in contexts: Australian perspectives and issues*. St Leonards, NSW: Allen & Unwin.
- MARSTON, D., FUCHS, L. S. & DENO, S.L. (1986). Measuring pupil progress: A comparison of standardised achievement tests and curriculum-related measures. *Diagnostique*, 11, 77–90.
- MARSTON, D. & MAGNUSSON, D. (1988). Curriculum-based measurement: District level implementation. In J. GRADEN, J. ZINS, & M. CURTIS (Eds.), *Alternative education delivery systems: Enhancing instructional options for all students* (pp. 137–172). Washington, DC: National Association of School Psychologists.
- MARSTON, D.B. (1989). A curriculum-based measurement approach to assessing academic performance: What is it and why do it? In M.R. SHINN (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18–78). New York: The Guilford Press.
- 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. *The Journal of Special Education*, 26, 162–180.
- MEHRENS, W.A. & CLARIZIO, H.F. (1993). Curriculum-based measurement: Conceptual and psychometric considerations. *Psychology in the Schools*, 30, 241–254.
- PARKER, R., HASBROUK, J.E. & TINDAL, G. (1992). Greater validity for oral reading fluency: Can miscues help? *Journal of Special Education*, 25, 492–503.
- POTTER, M.L. & WAMRE, H.M. (1990). Curriculum-based measurement and developmental reading models: Opportunities for cross-validation. *Exceptional Children*, 57, 16–25.
- RADENCICH, M.C. (Ed.) (1994). *Adult literacy: A compendium of articles from the Journal of Reading*. Newark, DE: International Reading Association.
- RESCHLY, D.J., KICKLIGHTER, R. & MCKEE, P. (1988a). Recent placement litigation: Part II. Minority EMR overrepresentation: Comparison of Larry P. (1979, 1984, 1986) with Marshall (1984, 1985) and S-1 (1986). *School Psychology Review*, 27, 22–38.
- RESCHLY, D.J., KICKLIGHTER, R. & MCKEE, P. (1988b). Recent placement litigation: Part III. Analysis of differences in Larry P., Marshall and S-1 and implications for future practices. *School Psychology Review*, 27, 39–50.
- RODDEN-NORD, K. & SHINN, M.R. (1991). The range of reading skills within and across general education classrooms: Contributions to understanding special education for students with mild handicaps. *The Journal of Special Education*, 24, 441–453.
- SHINN, M.R. (1988). Development of curriculum-based local norms for use in special education decision-making. *School Psychology Review*, 17, 61–80.
- SHINN, M.R. (Ed.) (1989). *Curriculum-based measurement: Assessing special children*. New York: The Guilford Press.
- SHINN, M.R. & HUBBARD, D. (1992). Curriculum-based measurement and problem solving assessment: Basic procedures and outcomes. *Focus on Exceptional Children*, 24(5), 1–20.
- SHINN, M.R., KNUTSON, N., GOOD, R.H., TILLY, W.D. & COLLINS, V. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, 21, 459–479.
- STONER, G., CAREY, S.P., IKEDA, M.J. & SHINN, M.R. (1994). The utility of curriculum-based measurement for evaluating the effects of methylphenidate on academic performance. *Journal of Applied Behaviour Analysis*, 27, 101–113.
- SWAIN, K.D. & ALLINDER, R.M. (1996). The effects of repeated reading on two types of CBM: Computer maze and oral reading with second-grade students with learning disabilities. *Diagnostique*, 21, 51–66.
- TAYLOR, R.L., WILLITS, P.P. & RICHARDS, S.B. (1988). Curriculum-based assessment: Considerations and concerns. *Diagnostique*, 14, 14–21.
- 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(1), 3–10.
- TUCKER, J.A. (1985). Curriculum-based assessment: An introduction. *Exceptional Children*, 52, 199–204.

- TUCKER, J.A. (1987). Curriculum-based assessment is no fad. *The Collaborative Educator*, 1(4), 4-10.
- WHELDALL, K. (1996). *The Wheldall Assessment of Reading Passages (WARP): Experimental Edition*. Unpublished manuscript, Macquarie University Special Education Centre, Sydney.
- WHELDALL, K., CENTER, Y. & FREEMAN, L. (1993). Reading recovery in Sydney primary schools. *Australasian Journal of Special Education*, 17, 51-63.
- WHELDALL, K. & MADELAINE, A. (1997). Should we measure reading progress and if so how? Extrapolating the curriculum-based measurement model for monitoring low progress readers. *Special Education Perspectives*, 6, 29-35.

