

Curriculum-based Measures:
Development and perspectives

Stanley L. Deno
University of Minnesota

Please do not copy or quote without permission from the author.

178 Pillsbury Dr. S.E., Minneapolis, MN 55455

Abstract

Curriculum-based Measurement (CBM) is an approach to measuring the academic growth of individual students. The essential purpose of CBM is to aide teachers in evaluating the effectiveness of the instruction they are providing to individual students. Early research focused on testing the utility and effectiveness of CBM for increasing the achievement of special education for students with learning disabilities. Extensions of CBM research now address a broad range of educational issues in both special and general education with different populations and in new curriculum domains. The present paper provides a brief history of CBM clarifying its relationship to curriculum-based assessment, a summary of the primary features of CBM, and a rationale for its design. Extensions of CBM to new research domains are identified, and a perspective is provided on the place of CBM in the broader context of educational assessment.

Curriculum-based Measurement (CBM) (Deno, 1985) is an approach to measuring the academic growth of individual students. The essential purpose of CBM has always been to aide teachers in evaluating the effectiveness of the instruction they are providing to individual students. Having said that, however, it is important to emphasize that much excellent research and development on CBM has extended it to educational decisions well beyond those for which it was originally created. Early work on improving the effectiveness of special education for students with learning disabilities has been expanded to screening and identification of students at risk of academic failure, to developing school wide accountability systems, to addressing the problem of disproportionate representation, to evaluating growth in early childhood, to assessing attainments in content area learning, to measuring literacy in students who are deaf, to assessing students who are English language learners (ELL), and to predicting success on high stakes assessments. What follows in this paper is a brief history of the development of CBM and some reflections on current efforts to use CBM to address a variety of educational problems.

Development of CBM

CBM originated in the Data-based Program Modification (DBPM) model described by Deno & Mirkin, (1977). That model outlined how a variety of progress monitoring data could be used to make educational programming decisions for students in special education. The DBPM model was designed for special education resource teachers to use in improving their interventions with students who were struggling

academically. While the model showed how the data could be used to make the types of decisions made for students in special education, its validity as an approach for improving special education had not been empirically validated. To explore the validity of DBPM, an empirical research and development program was conducted for six years through the federally funded of the University of Minnesota Institute for Research on Learning Disabilities (IRLD). The ultimate question pursued through that research was whether a formative evaluation system could be developed that teachers could use to improve their effectiveness in teaching students with academic disabilities. Ultimately, a comparative study demonstrated that teacher's were more effective using such a model (Fuchs et al,1984).

In the course of conducting the CBM research program through the IRLD we empirically developed a set of generic progress monitoring procedures that met conventional reliability and validity criteria in the areas of reading, spelling, and written expression. Three key questions were addressed in developing the CBM procedures: First, "What are the outcome tasks on which performance should be measured?" ("What to measure"), 2) "How must the measurement activities be structured to produce technically adequate data?" ("How to measure"), and 3) Can the data used to improve educational programs?" ("How to use"). The questions were answered through systematic examination of three key issues relevant to each – the technical adequacy of the measures, the treatment validity or utility of the measures, and the logistical feasibility of the measures. The framework for developing the measures has been specified elsewhere and will not be included here (Deno & Fuchs, 1987). The results of the research on progress monitoring led to the development of an assessment approach

typically referred to as CBM (Deno, 1985). Extensive field applications of research illustrating “what to measure” and “how to measure” and “how to use” the data have occurred are described in a variety of publications (cf., Fuchs, et. al., 1988; Shinn, 1989; Shinn, 1998). In summary, however, that research resulted in basic skills measures that are now widely used to improve educational decisions in a variety of contexts.

CBM and CBA

The term “Curriculum-Based Assessment” (CBA) became very popular in the field of Special Education with the publication of a special issue of Exceptional Children on that topic in 1985 (Tucker, 1985). In that issue, Tucker described CBA to as a practice that had existed for a very long time - the practice of using what is to be learned as the basis for assessing what has been learned. While his description is appealing, it doesn't clearly distinguish CBA from traditional psychometric test construction where a table of specifications is used to define the content domains of a test and the tests are then designed to test for whether that intended content has been learned. Four salient differences between CBA and traditional psychometric testing can be identified, however: First, in CBA, the very curriculum materials that serve as the media for instruction are used as the test stimuli, second direct observation and recording of student performance in response to selected curriculum materials is emphasized as a basis for collecting the information used to make assessment decisions; third, interobserver agreement is the primary technique used to establish the reliability of information collected through CBA; and fourth, social validity is typically the basis for justifying the use of information gathered through CBA. Given these emphases, it is common for CBA proponents to argue that the information gathered from student performance in the

curriculum more adequately reflect the real goals of instruction in the classroom than most standardized achievement tests, since the assessment information obtained through CBA relates more directly to what is being taught and, further, that the content and materials of daily instruction are a fairer and firmer basis for making judgments about student learning.

CBM as distinct from CBA. Since the focus in this paper is on CBM, distinguishing between CBM & CBA is necessary. The term “assessment” as used in CBA is a very broad term that refers to information gathered for purposes of decision-making. Thus, curriculum-based assessment is often used to refer to any information gathering practices that may occur when gathering information about student performance in the curriculum. These practices can include scoring a student’s worksheets to determine the percentage of questions answered correctly on a worksheet; doing an error analysis of a student’s oral reading from instructional text, or establishing “mastery” of a new skill based on performance on an end of unit test. In CBA, typically, different assessment information is collected for different decisions. A variety of different, but related, approaches to CBA are represented in the current literature (e.g., Howell, et al 1993; Bigge, 1988; Idol, Nevin, & Paolucci-Whitcomb, 1986; and Shinn, 1989).

“Measurement” rather than “assessment”. From the perspective provided here, Curriculum-Based Measurement (CBM) is a separate and distinct subset of CBA procedures. CBM refers to a specific set of procedures for measuring student growth in basic skills that were developed at the University of Minnesota through the Institute for Research on Learning Disabilities (Deno, 1985). The procedures were developed as part

of a larger program of research directed toward designing a practically feasible and effective formative evaluation system that special education teachers could use to build more effective instructional programs for their students. As part of that formative evaluation system it was necessary to create a simple, reliable, and valid set of measurement procedures that teachers could use to frequently and repeatedly measure the growth of their students in the basic skills of reading spelling and written expression. As with CBA, the measurement procedures of CBM become “curriculum-based” when the measurements are used within the context of the local school’s curriculum.

CBM and “General Outcome Measurement”

As continued development of CBM has occurred, evidence has been generated leading to the conclusion that the generic measurement procedures of CBM can provide technically adequate, instructionally relevant data using stimulus materials that have been drawn from sources other than a school’s curriculum (Fuchs & Deno, 1994). For that reason, the terms “General Outcome Measurement GOM” (Fuchs & Deno, 1994) and Dynamic Indicators of Basic Skills (DIBS)” (Shinn, 1995,1998) have been coined to refer to the generic measurement procedures used with stimulus materials that are not drawn from the curriculum. This “uncoupling” of CBM with necessity for using the local school’s curriculum has made it increasingly possible in research and practice to capitalize on using standardized stimulus materials without loss of the relevance of CBM for making every day instructional programming decisions. Further, it has enabled extensions of CBM to areas of skill development where schools do not always have a curriculum (e.g., secondary reading and written expression – Espin & Deno, 1993; Espin, et. al. 1999; early literacy – Kaminski & Good, 1996; and English Language Learning –

Baker & Good, 1995). It has also facilitated development of computer-based applications (Fuchs, et. al, 1993), enabled aggregation of data across schools to make district level evaluation decisions (Marston & Magnusson, 1998), use as a component of effective classroom intervention packages (Fuchs, et al, 1997), opened new avenues to assessing reading and writing in students who are deaf and hard of hearing (Chen, 2002; Devenow, 2002). Further examples of the development and extensions of the generic CBM procedures for measurement are illustrated through articles in this special issue (Busch et al; Lembke &

An Example o f CBM

A distinctive characteristic of CBM when used to improve individual student programs is the individual student progress graph illustrating the responsiveness of a student to various program modifications made by the teacher. In Figure 1 is an illustration of the results of using CBM procedures with a student in reading over the course of a single school year. What we see in the figure is the results of repeated sampling of student performance in reading different passages from the same book throughout the year. On the graph we see plotted the number of words read aloud correctly in one minute from those passages. In addition, we see how that performance changes in relation to changes made in that student's program.. The graph is important because it provides the user of CBM with the framework used to evaluate the effects of efforts to improve the student's rate of growth. It is this continuous evaluation framework that was tested by Fuchs, Deno and Mirkin (1984) and found effective in accelerating that growth.

Insert Figure 1 here

Characteristics of CBM

Repeated measurement on a single task. The generic measurement procedures used in CBM and GOM are based on obtaining repeated samples of student performance on equivalent forms of the same tasks across time. Changes in performance on this task are then interpreted to reflect generalizable change in a student's proficiency at that task. The procedures are analogous to what occurs when we measure the change in a child's height and weight using a scale or ruler. The concept is simple, but it is uncharacteristic of education. From the outset CBM development has attempted to create a system where teachers are able to focus clearly on the target of their instruction. The assumption was that successful intervention required that teachers receive clear and unambiguous feedback regarding the general effects of their instructional efforts. If teachers are either uncertain about the overall effects of their efforts, or believe they have been successful simply because a student learns the specific content that has been taught, their efforts to improve growth will be unsuccessful. The uncertainty present in the former case can stem from the fact that teachers do not have "vital sign" indicators for learning, such as pulse rate and temperature, that they we can use to monitor the effects of their instruction in basic skill outcomes. In many respects teachers must operate like early pilots who had to resort to feel; that is, to "flying by the seat of their pants" because instruments to

indicate aircraft altitude and attitude had not yet been developed. Unfortunately, without instrumentation it was possible for pilots to believe they were flying straight and level when they were headed directly toward the ground. It is also possible for teachers to believe that they have been successful because a student has learned what has been taught. The misfortune in this case is that it is possible to successfully teach something that might not contribute to developing the overall proficiency for which the curriculum is designed. The problem is similar to what occurs when taking a golf lesson and learning to do what the instructor has taught, but then finding, much to our disappointment, that we play no better on the golf course. What we teach in an area like reading, and what students learn from that instruction, does not always contribute to general improvements in reading. The measurement tasks of CBM have been empirically selected because they reflect whether the instruction we are providing does, in fact, result in improvement in general reading outcomes.

Empirically selected tasks. In developing CBM procedures, a two-part strategy is used to identify those tasks that teachers can reasonably use to evaluate their instruction. The first part of the strategy ---initial task selection ---is based on research using a criterion validity paradigm to select those tasks that would seem to be the best candidates for repeated performance measurement (Deno, 1985; Marston, 1989; Fuchs, et al, 1988). The second part of the task selection strategy is to test the instructional utility of the measures through evaluating the student achievement effects of teachers using the CBM data to make instructional evaluation decisions (Fuchs, Deno, and Mirkin, 1984; Fuchs, 1989).

Reliability. A related consideration in empirical task selection relates to the fact

that CBM data are used to make important instructional intervention decisions. For that reason, like criterion validity, the tasks selected for use in CBM have always been ones for which reliable measures could be constructed (Marston, 1989). It should be noted that establishing the reliability of CBM has always included not only interobserver agreement, but test/retest and alternate form reliabilities, as well. The latter reliability is particularly important since multiple forms are used in CBM.

Economical and efficient. A number of important additionally important characteristics used in developing CBM procedures relate to the need for them to be logistically feasible within the context of ongoing instruction.

Time efficient—Since frequent, repeated measurement are required for growth measurement and evaluation, CBM tasks must be of short duration.

Multiple forms —Each repeated measurement of CBM must be in response to a stimulus task that is unfamiliar to the student so that any increase in performance represents real growth in general proficiency rather than the effects of practice. Thus any task to be used must be for which it is simple to create many equivalent forms.

Inexpensive—Since many forms must be made available for teachers to use frequently, the task has to be one that would not require expensive production of materials.

Easy to teach—Since it is likely that many teachers, paraprofessionals and students will administer the measures, the task must be one that could be easily taught.

Issues in implementing CBM. Establishing parameters such as these in task selection for CBM has always been important because it delimits the range and variety of tasks included in any search for valid indicators of basic proficiency. In addition, specifying the characteristics of a practically feasible task on which to do frequent, repeated measurement enables the developer to focus our criterion validity research on only those tasks that reasonably could be part of a classroom- based, ongoing formative evaluation system. Unfortunately these reasons for limiting task selection have not always been fully understood or appreciated by many who have first encountered CBM. That failure can lead both potential users and developers of alternative measures to increase the complexity and scope of measures intended to assess curriculum outcomes. The result can be impractical measures that cannot be used as part of routine classroom instruction.

Paradoxically, the fact that CBM procedures do not require tasks for measurement that seem sufficiently complex can also mislead critics into believing that the CBM measures are invalid (Shinn, 1998). A good illustration of the problem is in the area of reading where the evidence has been developed that, when structured properly, “reading aloud from text” can be used to develop a global indicator of reading proficiency (Deno, Mirkin, and Chiang, 1982). The major criticism of measuring reading by having students read aloud from connected discourse is that such a task does not reflect a student’s comprehension of text. On technical grounds this criticism is invalid (Fuchs, et al, 1988; Good & Jefferson, 1998). The criterion validity research on using this task in reading measurement provides a solid empirical basis for concluding that the number words read aloud correctly from text in a one minute time sample is a good indication of a student’s

general reading proficiency. CBM reading scores relate sensibly to standardized achievement test scores, to students' age and grade, to teacher's judgments of reading proficiency, and to their placements in regular, compensatory and special education programs. Despite this critics will argue that our CBMs in reading should include a "direct" measure of comprehension"; such as, answering comprehension questions or retelling the story that has been read.

While it is possible to argue on empirical grounds that reading aloud from text indexes comprehension better than most so called "direct measures" (Fuchs, Fuchs, and Maxwell, 1988), it is more to the present point to clarify that tasks such as "answering comprehension questions" or "retelling the story" do not meet the requirements established for the CBM that were outlined above. To use either task would a) consume far too much time to be used for repeated measurement in CBM (students would have to read fairly lengthy passages so that question asking or story retelling would be possible), b) cost too much in the development of multiple equivalent forms, and c) as in the case of story retell, be difficult to teach others to score reliably. Thus, while these tasks have been used as criterion measures in CBM task selection, they must be excluded as candidates for repeated measurement on other important grounds.

As CBM developers have painfully learned, however, neither empirically nor technologically valid reasons are enough to persuade many people. In a study of the barriers to successful use of CBM (Yell, et al, 1992) "face validity" issues stood out as among the most important concerns for teachers. In their survey, Yell, et al, also found interesting differences between teachers and their administrators. Teacher focused on the immediate impact of using CBM on a frequent basis and expressed concern about the

additional time required in doing CBM. In fact, three of the five frequently identified barriers by teachers refer to time associated problems – this, despite the efficient nature of the measures. The administrators view of problems associated with implementing CBM were quite different from the teachers. The emphasis in their responses was that it was difficult to develop effective teacher use of the CBM procedures. Three of the five most frequently identified barriers by administrators addressed difficulties related to a lack of teacher resourcefulness in using the CBM data responsively to modify and evaluate their instruction. Of interest is the fact that the single most frequently identified barrier from the administrator’s perspective was the “natural resistance” that occurred when any change in practice was required of school personnel.

Reflections on CBM in the Broader Context of Assessment

The results of the CBM research program have provided a basis for developing standardized measurement procedures that can be used to formatively evaluate the effects of modifications in the instructional programs for individual students. Indeed, the research conducted on the student achievement effects of special education teachers using these procedures provides a basis for concluding that instructional effectiveness can be improved through the use of CBM in formative evaluation. At the same time, the CBM procedures have been used to data-base the full range of intervention decisions that are made for students who are academically “at risk”. In addition, CBM/GOM.DIBS are being used to addressing the problem of disproportionate representation (Minneapolis Public Schools, 2001), in a Problem Solving Model that emphasizes prereferral intervention evaluation (Shinn, 1995; Tilly & Grimes, 1998), to appraise growth in early childhood (Kaminski & Good, 1996)., to assessing attainments in content area learning

(Espin & Foegen, 1996), and to predicting success on high stakes assessments (e.g., Deno, et al, 2002; Good, Kameenui & Simmons, 2001).

Developments in using CBM procedures have accelerated quite dramatically in the past five years. Textbooks now routinely include descriptions of how CBM is used in both assessment and remediation (Mercer, 1997; Spinelli, 2002; Taylor, 2000; Henley, Ramsey, & Algozzine, 2002). Dissemination is now extensive, and much of that dissemination is likely due to the functional utility of the measures. It is also likely that the generic nature of the procedures has allowed a wide range of potential users and developers to feel “ownership” over both the procedures and the data collected. This sets apart the standardized procedures of CBM from the most standardized tests that are the commercial property of the test developers and the companies that publish those tests.

It will be interesting to track the relative use of growth measures like CBM and status measures like commercial standardized tests.. Very little work has been done in the private sector to develop progress monitoring systems. Why that should be is unclear, but the reason probably lies in the fact that development of educational and psychological measurement in the United States has been directed toward discriminating *between* individuals for purposes of classification. Tests used for that purpose are designed to describe differences between individuals rather within an individual across time.

Differences between individuals are important when the primary function of assessment is to sort individuals into groups for making selection decisions rather than to examine individual growth. Those of us in working in school programs are very much aware that assessments commonly are conducted to classify students as eligible for alternative programs like special education, Title I, and gifted education. In all of these

cases, the decision to be made has rested on distinguishing the relative achievements or accomplishments of a subgroup of students within the general student population. Since the economic and social consequences of these decisions are potentially very important, it is not surprising that responsible decision makers would seek assessment procedures that discriminate and quantify differences between individuals as justification for these decisions.

Interest is now increasing in the idea that we should examine individual performance to ascertain attainment of “standards.” Criterion performance on particular tasks is gaining prominence in the view of decision makers. Important also in this shift to alternative approaches to performance assessment is not only the increased emphasis on criterion performance, but also on the nature of the tasks selected for assessment purposes. “Authenticity” has become the prime characteristic to be embraced when tasks are selected, and, for that reason, “face validity” has now become paramount in task selection. Indeed, the argument is that authenticity and face validity can take the place of the more traditional reliability and validity criteria of psychometrics.

If we are interested in developing CBM procedures for continued use in student progress monitoring, I think we must see the recommendations associated with alternative approaches to assessment as helpful. Contained in those recommendations is an emphasis on individual attainment that is at the basis of progress monitoring. Discriminating growth relative to a performance standard is an important shift in attention away from the emphasis on making distinctions between individuals. At the same time we should not be sanguine about the possibility that the focus will now become individual growth rather than sorting and classifying students. Indeed, those of

us concerned with the education and habilitation of people with disabilities have already seen that the emphasis on attaining performance standards has resulted in the tendency to exclude such persons from the assessment process. A second concern is that the race to develop alternatives has resulted in expectations far exceeding reality. Establishing authenticity and instructional utility as characteristics for assessment are admirable ideals, but just as developing a cure for cancer requires more than specifying the goal, developing assessment procedures with particular characteristics requires more than asserting their importance. Contrary to assumptions currently made by advocates of “authentic assessment”, the technical knowledge required for accomplishing our goals is neither available nor unnecessary. Any reading of the current literature on the results of efforts to develop and use new alternative approaches to assessment reveals that the effort is fraught with difficulty.

Many years ago, Jerome Bruner argued that achievements in developing technology that increase our powers of observation is at the basis of most of our greatest scientific achievements. If that is so, then the development of improved procedures for assessing individual growth may well result in breakthroughs that increase our knowledge of human development and our successes in optimizing that development. Most certainly, breakthroughs in assessment technology expand our knowledge in the long run will result from the types of intense research and development efforts presented in this special issue rather than from engaging in the politics of education..

References

- Baker, S. K., & Good, R. H. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: A validation study with second-grade students. School Psychology Review, 24, 561-578.
- Bigge, J. (1988) Curriculum-based instruction. Mountain View, CA:Mayfield Publishing Co.
- Chen, Yi-ching (2002) Assessment of reading and writing samples of deaf and hard of hearing students by Curriculum-based Measurements. Unpublished doctoral dissertation. Minneapolis: University of Minnesota
- Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.
- Deno, S.L. & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for databased special education problem solving. Focus on Exceptional Children, 19 (8), 1 - 15.
- Deno, S. L. & Mirkin, P. K. (1977) Data-based Program Modification: A Manual. Reston VA: Council for Exceptional Children.
- Deno, S. L., Reschly-Anderson, A., Lembke, E., Zorka, H. Callender, S. (2002) A Model for School wide Implementation: A Case Example. Presentation at the National Association of School Psychology Annual Meeting. Chicago.
- Devenow, P. S. (2002) A study of the CBM maze procedure as a measure of reading with deaf and hard of hearing students. Unpublished doctoral dissertation. Minneapolis: University of Minnesota
- Espin, C.A., & Deno, S.L. (1993). Content-specific and general reading disabilities of secondary-level students: Identification and educational relevance. Journal of Special Education 27, 321-337.
- Espin, C.A., & Foegen, A. (1996). Validity of three general outcome measures for predicting secondary student performance on content-area tasks. Exceptional Children, 62, 497-514.
- Espin, C.A., Scierka, B.J., Skare, S., & Halverson, N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary students. Reading and Writing Quarterly, 15, 5-27.

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

Fuchs, L. S., Fuchs, D., & Hamlett, C.L. (1989) 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., & Maxwell, L. (1988). The validity of informal reading comprehension measures. Remedial and Special Education, 9, 20-28.

Shinn, M. R. Ed. (1998), Advanced applications of curriculum-based Measurement (pp. 61-88). New York: Guilford.

Good , R. & Jefferson, G. (1998) Contemporary perspectives on curriculum-based measurement validity. In M.R. Shinn (Ed) Advanced applications of Curriculum-based Measurement. New York: The Guilford Press. (pp 61-88).

Good, R. H.III, Simmons, D.C.& Kameenui, E. J. (2001) The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high stakes outcomes. Scientific Studies of Reading, 5(3), 257-288.

Henley, M Ramsey, R. A., & Algozzine, R. F. (2002) Teaching Students with Mild Disabilities (4th Ed.). Boston: Allyn & bacon.

Idol, L., Nevin, A., & Paolucci-Whitcomb. (1986). Models of Curriculum-based Assessment. Rockville, MD: Aspen Publishers, Inc.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. School Psychology Review, 25, 215-227.

Marston, D. (1989) A curriculum-based approach to assessing academic performance: What it is and why do it. . In M. R. Shinn (Ed.), Curriculum-based measurement: Assessing Special Children, (pp 19-78). NY: Guilford Press

Mercer, C. D. (1997) Students with Learning Disabilities (5th Ed). Upper Saddle River, N.J.: Merrill/Prentice-Hall

Minneapolis Public Schools (2001) Report of the external review committee on the Minneapolis Problem Solving Model.

Shinn, M.R. Ed. (1989) Curriculum-based measurement: Assessing Special Children, (pp 1-17). NY: Guilford Press.

Shinn, M. (1995) Best practices in curriculum-based measurement and its use in a problem-solving model. . In J.Grimes and A. Thomas (Eds.), Best Practices in School Psychology III. (pp. 547-568). Silver Spring, MD: National Association of School Psychologists.

Shinn, M. R. Ed. (1998), Advanced applications of Curriculum-Based Measurement (pp. 61-88). New York: Guilford

Spinelli, C. (2002) Classroom assessment for students with special needs in inclusive settings. Upper Saddle River, NJ: Merrill/Prentice Hall

Taylor, Ronald L. (2000). Assessment of Special Students (5th Edition). Allyn & Bacon: Needham Heights, MA.

Tucker, J. (1985). Curriculum-based assessment: An introduction. Exceptional Children, 52(3) 266-276.

Yell, M., Deno, S. L., & Marston, D. (1992). Barriers to implementing curriculum-based measurement. Diagnostique, Fall, 18(1), 99.