Using Curriculum-Based Measurement to Evaluate Intervention Efficacy

Gary Stoner
Stanley E. Scarpati
Robin L. Phaneuf
John M. Hintze

SUMMARY. The use of Curriculum-Based Measurement for evaluating treatment efficacy is described and discussed. The basic methods and applications of Curriculum-Based Measurement are described, followed by examples of its application in two intervention program evaluation projects. The first project involved one individual student experiencing academic difficulties. The second project focused on evaluating a reading instruction program for a group of students. In both examples, the use of Curriculum-Based Measurement contributed significantly to monitoring student progress in response to educational program variables and making treatment evaluation decisions.

KEYWORDS. CBM, academic instruction, educational intervention

---

Gary Stoner, Stanley E. Scarpati, Robin L. Phaneuf, and John M. Hintze are affiliated with the University of Massachusetts at Amherst.

Address correspondence to: Gary Stoner, PhD, School Psychology Program, School of Education, University of Massachusetts at Amherst, Amherst, MA 01003-4150.

[Haworth co-indexing entry note]: "Using Curriculum-Based Measurement to Evaluate Intervention Efficacy." Stoner, Gary et al. Co-published simultaneously in Child & Family Behavior Therapy (The Haworth Press, Inc.) Vol. 24, No. 1/2, 2002, pp. 101-112; and: Behavior Psychology in the Schools: Innovations in Evaluation, Support, and Consultation (ed: James K. Luiselli, and Charles Diament) The Haworth Press, Inc., 2002, pp. 101-112. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: getinfo@haworthpressinc.com].

The environment for professional practice in today's public schools increasingly is one characterized, and sometimes dominated, by discussion of accountability, data-based decision making, and program effectiveness. Administrators, educators, parents, students, and other community members all are interested in a positive return on their varying forms of educational investment. Professional organizations across areas of education and psychology also emphasize these themes. For example, the National Association of School Psychologists has endorsed among its top priorities for professional training and practice, data-based decision making and accountability, effective instruction, prevention, wellness promotion, and program evaluation (Ysseldyke et al., 1997). Similarly, the American Psychological Association emphasizes evidence-based intervention strategies in professional training and practice (Kratochwill & Stoiber, 2000; Stoiber & Kratochwill, 2000), and the National Council for the Accreditation of Teacher Education (National Council for the Accreditation of Teacher Education, 2001) standards for personnel preparation emphasize knowledge and skill development with a focus on student learning. As such, instructional and curricular programs of all types are the focus of increasing scrutiny, with an eye toward the provision of continued support for what works to improve child/student outcomes and the discontinuation of support for what does not work.

Evaluations of what does and does not work are particularly important for students experiencing behavior and achievement problems, and who may be characterized as difficult-to-manage and/or difficult-to-teach. Continuation of ineffective interventions places these students further behind their typically developing peers, and at increasingly greater risk for adjustment problems, school dropout, referral to special education, and a host of problematic life outcomes (Patterson, Reid, & Dishion, 1992). Conversely, the promotion of positive educational and life outcomes for these students is, at least in part, dependent on the design, implementation, and evaluation of effective interventions. As Ikeda, Tilly, Stumme, Volmer, and Allison (1996) have discussed with regard to district-wide innovations in school-based problem solving, one critical foundation of such efforts is a commitment to measuring student performance frequently and changing programs when students are not progressing.

One important tool useful for systematically evaluating interventions through measurement of student performance is Curriculum-Based Measurement (CBM). Developed over the past two decades (Deno, 1992; Shinn, 1998), CBM is user friendly (Shinn, 1989), has founda-

tions in behavioral assessment (Shinn, 1998), is grounded in a problem solving approach to educational practice (Deno, 1995), and is useful for making program evaluation decisions both for individual students as well as groups of students (Deno, 1986).

In the following sections we provide a basic description of CBM, followed by two examples of its use in consultation to evaluate intervention efficacy. The first example focuses on interventions for an individual student achievement problem. The second example describes a program evaluation project examining a reading instruction program for a group of students identified as at risk for or having reading problems.

WHAT IS CURRICULUM-BASED MEASUREMENT?

Grounded in behavioral assessment, Curriculum-Based Measurement (CBM) was developed to both provide a technology for systematic, formative evaluation of student academic outcomes in the basic skill areas of reading, spelling, writing, and math, and to support intervention effectiveness evaluations using single-case study designs (Deno, Mirkin, & Chiang, 1982). Primary CBM data are derived from brief (1- to 3-min) fluency measures of student performance. These measures are content valid in that the materials used to evaluate outcomes are sampled directly from the student's curriculum. Also, the measures assess important and socially valid terminal behaviors, such as the number of words read correctly, number of correct letter sequences written, number of correct math problems or correct digits written.

A goal of CBM is to provide teachers with assessment information that can be used to plan instructional programs and to evaluate overall student growth. This goal is facilitated by CBM yielding rate-based measures of student performance, such as words read correct per minute (WRC). WRC then often is translated into slope of progress over time, such as 2 WRC gain per week over a 4-week time frame. Another distinctive feature of CBM is that an extensive body of research has accumulated to support the technical adequacy of the principal measures from both behavioral and traditional psychometric perspectives (Fuchs & Deno, 1991; Good & Jefferson, 1998). Further, the standardized procedures for conducting CBM probes are designed for simple, low-cost, repeated administration (Knutson & Shinn, 1991).

A systematic program of research was initiated in the late 1970s to investigate the technical adequacy of the standardized academic tasks

used in CBM (Marston, 1989). Since then, an extensive body of empirical evidence has been accumulated that supports the reliability and validity of CBM procedures for educational decision-making. For example, in the CBM reading task, the number of words read correctly has been validated as a reliable and accurate measure of students' general reading skills, including reading comprehension (Shinn, Good, Knutson, Tilly, & Collins, 1992). A summary of the standardized tasks and scoring procedures along with references to several studies documenting the technical adequacy of CBM reading, math, spelling and written expression measures are presented in Table 1. A number of researchers have demonstrated that CBM data are sensitive to changes in student performance as a result of various instructional interventions (Deno, Mirkin, & Chiang, 1982; Marston, Fuchs, & Deno, 1986), and as a result of interventions to address social behavior problems in classrooms. For example, CBM measures have been used to evaluate the effects of computer-assisted instruction (Fuchs, 1988), classwide peer-tutoring (DuPaul & Henningson, 1993), goal-setting strategies on students' academic achievement (Fuchs, Fuchs, & Hamlett, 1989), and stimulant medication on the academic performance of children presenting with disruptive behavior disorders (Stoner, Carey, Ikeda, & Shinn, 1994).

TABLE 1. CBM measures of reading, math, spelling and written expression: Standardized tasks, scoring procedures and technical adequacy research

Academic Area	Task	Scoring	Technical Adequacy Evidence
Reading	Students read passages aloud for 1 minute.	Number of words read correctly. Number of errors.	Deno, Mirkin, & Chiang (1982); Fuchs, Fuchs, & Maxwell (1988)
Spelling	Students write words dictated orally for 2 minutes.	Number of correct letter sequences. Number of words spelled correctly.	Deno, Marston, Mirkin, Lowry, Sindelar, & Jenkins (1982); Marston, Lowry, Deno, & Mirkin (1981)
Written Expression	After being given a story starter or topic sentence, students write a story for 3 minutes.	Number of words written. Number of words spelled correctly. Number of correct word sequences.	Deno, Marston, & Mirkin (1982)
Math	Students write answers to written computation problems for 2-5 minutes.	Number of correct digits.	Fuchs & Fuchs (1987); Marston, Fuchs, & Deno (1986)

USING CBM TO EVALUATE A READING INTERVENTION PROGRAM FOR ONE STUDENT

Michael was a first-grade student referred by his teacher with concerns centering around his lack of progress in reading. As reported by the teacher, while typical students in the class were being instructed in the Grade-1 reader, Michael was being taught in Primer Level books (e.g., kindergarten level picture books) and still was experiencing frustration with respect to word recognition and reading fluency.

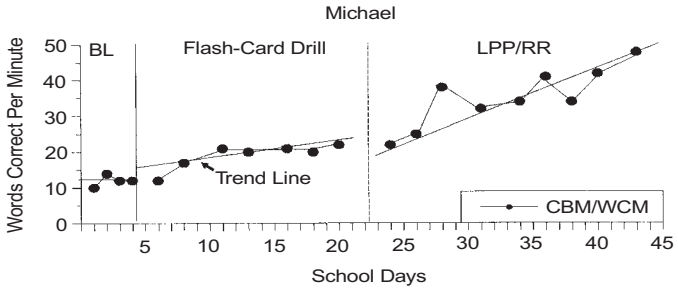
Upon initial assessment, Michael was asked to read isolated words presented on individual flash cards. From a pool of approximately 500 words, Michael was able to accurately read only 75 of the words presented. In addition to individual words, Michael also was asked to read selected passages from the material in which he was being instructed (i.e., Primer Level). When asked to do so, Michael on average was able to read only about 12 words correct per minute. School-based CBM norms indicated that typical first-grade students at that time of the year were able to read approximately 37 words correct per minute in Grade-1 material. Overall, results of the reading assessment indicated that, relative to his classmates, Michael was experiencing considerable difficulty in his beginning reading skills and was at risk for further reading problems.

Based on the results of the initial assessment, an intervention strategy was proposed that focused on building and strengthening Michael's word recognition and sight-word vocabulary skills. Specifically, a flash-card drill strategy was designed that contained both words that Michael had mastered and words that he had yet to learn (Gickling & Havertape, 1981). Mastered words were incorporated into the intervention to enhance Michael's sense of competence through success, as well as to foster his motivation to tackle new unlearned words. As presented, the flash-card drill and practice intervention contained all essential elements of effective instruction including modeling of accurate word reading by the teacher, prompting, and frequent opportunities to respond to instructional material. When words were pronounced incorrectly, the teacher provided the correct pronunciation, asked Michael to repeat the correct pronunciation, and then use the word correctly in a sentence. Furthermore, when Michael correctly read a previously unknown word five times in a row the word was considered mastered. When this happened, the newly mastered word was replaced by a new unmastered word in the flash-card drill packet.

The effect of the flash-card drill strategy on CBM oral reading fluency skills is presented in Figure 1. As can be seen, prior to any inter-

vention Michael read on average about 12 words correct per minute in Grade-1 Level material. Initiation of the flash-card drill strategy improved Michael's oral reading fluency to about 19 words read correct per minute. This amounted to a gain of 7 words correct per minute over a two-week period, or about 3.5 words correct gain per week. Examination of the trend line during this first phase of intervention suggests, however, that although Michael was improving in his oral reading fluency skills at a rate of improvement that was normally expected of first-grade students, continued growth at this rate would not ameliorate the discrepancy between his current level of performance and that of his same age peers. Figure 1 data also indicate that after some initial improvement from days 5 to 11 in Michael's oral reading fluency skills, the slope of progress tapered off with negligible improvement observed over the following two weeks. Left unattended, the discrepancy between Michael's oral reading fluency skills and those of his peers would continue to grow. This prediction led to a decision to change the nature of the intervention and incorporate strategies that explicitly targeted oral reading fluency skills in connected text. In doing so, two intervention strategies were combined: (a) listening passage preview (LPP), and (b) repeated readings (RR) (Daly, Martens, Hamler, Dool, & Eckert, 1999). During this phase of intervention, the teacher first read a selected reading passage at a comfortable reading pace, and instructed Michael to follow along with his finger and read the words to himself (i.e., LPP).

FIGURE 1. Curriculum-based measurement oral reading fluency data for Michael as a function of baseline and intervention conditions.



Note. BL = Baseline; LPP/RR = Listening Passage Preview/Repeated Reading. Portions of this figure are reproduced from Daly, Hintze, and Hamler (2000) with permission from John Wiley & Sons, Inc.

After this, Michael was asked to read the same passage three times in succession with the total amount of time required to finish the passage noted each time. The time it took to read the passage each time was graphed and combined with verbal praise and reinforcement for lowering the time it took to read the passage with each successive attempt. The results of this second phase of the intervention on CBM oral reading fluency skills can also be seen on Figure 1. With the initiation of LPP/RR, Michael's rate of improvement increased significantly. On average, Michael's overall reading rate increased from 19 words correct per minute during the first phase of intervention to 35 words read correct during the second phase of intervention. More importantly, however, the slope of Michael's growth was substantially greater as a result of LPP/RR. Specifically, Michael's rate of growth increased from about 3.5 words correct per week during the first intervention phase, to about 7 words correct per week during the second intervention phase. Such growth is nearly triple that of what can otherwise be normally expected of typically developing first-grade students. In comparing the trend lines across the two phases it is easy to see that the fluency based intervention was demonstrably better than the flash-card drill alone and served Michael well in helping to close the gap between his reading performance and that of his peers. Indeed, at the ninth week of intervention Michael was reading upwards of 50 words correct per minute in the Grade-1 level reader, a level commensurate with typical peer performance at that time of the school year in his school.

USING CBM TO EVALUATE A GROUP READING INTERVENTION PROGRAM

Educators often are faced with the need to determine the effectiveness of instruction, that is, to simply answer questions about the efficacy of a particular curriculum or teaching strategy for a group of students, for example, at the program, classroom, or building level. The purpose of the project presented here was to evaluate a experimental reading instruction program (i.e., one with unproven efficacy) aimed at improving the reading skills of students with reading disabilities and/or students at risk for developing such problems. The program had been in place for a number of years, yet few data had been collected to speak to its effectiveness. Our primary goal was to determine the efficacy of the instruction and the progress of the students receiving instruction. The focus of the evaluation was on the progress in reading skills made by ten

grade 3, 4, and 5 student participants, randomly selected from the total of twenty students in the experimental program. The same outcome measures were used for all children, and consisted of repeated curriculum-based measures of reading–specifically oral reading fluency. For each student in the evaluation, three oral reading fluency probes were administered at one sitting, on a weekly basis throughout the course of the evaluation, and the median score was taken as representative of that week's student performance.

CBM was selected as the tool for evaluating three aspects of student performance. First, CBM was used to determine the growth trajectories of students in the specialized reading group as indicated by rates of oral reading fluency. CBM data were used to determine the individual rates of progress as measured using time-series slopes of progress over a 10-12 week period of time. The slopes were described as words per week gain over time. Using both professionally agreed upon criterion for expected progress (i.e., 2-3 words per week gain in progress monitoring materials; see Fuchs et al., 1993), and projections of rates of gain necessary to approximate typical peer achievement, CBM data were used to determine how many children in the sample could be characterized as being successful or not successful in reading. Second, CBM was used to compare growth trajectories of students in the experimental group to students receiving instruction in general education (Shinn, 1988). The majority of children in an effective remedial reading program should experience average to above average gains via program participation. Comparison of experimental group students to students receiving instruction in general education provides a means of evaluating whether students in the experimental group are making gains over and above their peers. Finally, CBM was used to predict achievement trajectories beyond the time frame of the evaluation.

Data for five of the ten children evaluated are summarized in Table 2; the other five all entered the experimental program reading at skill levels above the 25th percentile relative to district norms, and as such were considered inappropriate candidates for continued program participation by our evaluation team. Outcomes of the evaluation generally indicated students in the experimental group most in need of remedial reading instruction typically were making less than average progress. Those participants least in need of remedial reading instruction typically were making average to above average progress. There were two exceptions to these findings in that one student with poor beginning level skills was found to be making above average progress, and one student with reasonably good beginning level skills was found to be making very little progress in reading. In general, the findings suggest a

phenomenon of "the rich getting richer, and the poor getting poorer"–certainly not the intent of a specialized reading instruction program for low performing students. In addition, but beyond the scope of this paper, the data called into question the methods for selecting students for participation in the specialized program.

The data in Table 2 for two students, FG and JD, represent two of the major findings of the evaluation. During the course of the evaluation, FG was observed to be making progress in oral reading fluency at the rate of 2.7 words gain per week over the course of eleven weeks. Given this rate of progress and FG's beginning percentile rank, FG should be considered a student who is unsuccessful in reading skill development. FG needs to be making much greater progress in order to improve his relative standing, in the same manner as Michael in our first case study needed an additional intervention. By comparison, JD was observed to be making progress in oral reading fluency at the rate of 0.0 words per week gain. Given this lack of progress over three months time, and JD's beginning percentile rank, JD should be considered a student who is not successful in reading skill development. In order to improve her relative

TABLE 2. Reading achievement and progress data for experimental reading program students whose entry level reading skills were below the 25th percentile of local normative data.

Student; grade in school	Percentile rank at program entry	Words read correct gain per week	Reading progress relative to expected progress	General conclusion
NP; 4th grade boy	7th	3.5	Above average	Making above average progress; should be reintegrated into general education reading instruction
FG; 4th grade boy	1st	2.7	Average	Making average progress; needs to make above average gains, given such low entry skills
KH; 5th grade girl	9th	1.45	Below average	Making poor progress; needs intensive, more effective instruction
TS; 3rd grade girl	7th	1.25	Below average	Making poor progress; needs intensive, more effective instruction
JD; 3rd grade girl	6th	0.00	Below average	Making no progress; needs intensive, more effective instruction

standing, she needs to be making much better progress than she currently is.

The use of CBM in this program evaluation clarified issues of identification of students in need of specialized instruction and accountability procedures to ensure exiting of successful students and additional support for students still struggling despite remedial efforts. Specific design constraints limited the ability of the evaluation to further address the effectiveness of the experimental reading instruction program. The most legitimate evaluation of such a program would include at least the following components:

1. random assignment of identified children to experimental and control groups,
2. blinded, placebo controlled evaluations of treatment outcomes,
3. carefully examined and documented manipulations of the treatment components thought to be salient or powerful,
4. use of outcome measures meeting professional standards for technical adequacy and administered under standardized conditions.

However, despite the methodological limits placed on this evaluation, the CBM based approach yielded valid and useful data about the efficacy of the experimental reading program.

In summary, Curriculum-Based Measurement of basic academic skills represents a technically adequate and useful set of strategies for measuring and monitoring student achievement. Use of these strategies in monitoring student progress can contribute significantly to questions of intervention program efficacy at both the individual and group level, for interventions for achievement and behavior problems.

REFERENCES

Daly, E. J., III, Hintze, J. M., & Hamler, K. R. (2000). Improving practice by taking steps toward technological improvements in academic intervention in the new millennium. Psychology in the Schools, 37, 61-72.

Daly, E. J., III, Martens, B. K., Hamler, K. R., Dool, E. J., & Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. Journal of Applied Behavior Analysis, 32, 83-94.

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

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

Deno, S. L. (1995). School psychologists as problem solvers. In A. Thomas and J. Grimes (Eds.), Best Practices in School Psychology (3rd Ed.), pp. 471-484. Bethesda, MD: National Association of School Psychologists.

Deno, S. L., Marston, D., & Mirkin, P. K. (1982). Valid measurement procedures for continuous evaluation of written expression. Exceptional Children, 48, 368-371.

Deno, S. L., Marston, D., Mirkin, P. K., Lowry, L., Sindelar, P., & Jenkins, J. (1982). The use of standardized tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study. Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

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

DuPaul, G. J., & Henningson, P. N. (1993). Peer tutoring effects on the classroom performance of children with attention deficit hyperactivity disorder. School Psychology Review, 22, 134-143.

Fuchs, L. S. (1988). Effects of computer-managed instruction on teachers' implementation of systematic monitoring programs and student achievement. Journal of Educational Research, 81, 294-304.

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

Fuchs, L. S., Fuchs, D., & Hamlett, C. (1989). Effects of alternative goal structures within curriculum-based measurement. Exceptional Children, 55, 429-438.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much 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.

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

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

Ikeda, M. J., Tilly III, W. D., Stumme, J., Volmer, T., & Allison, R. (1996). Agency-wide implementation of problem solving consultation: Foundations, current implementation, and future directions. School Psychology Quarterly, 12, 28-39.

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

Kratochwill, T. R., & Stoiber, K. C. (2000). Empirically supported interventions and school psychology: Conceptual and practical issues–Part II. School Psychology Quarterly, 15, 233-253.

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

Marston, D., Fuchs, L. S., & Deno, S. L. (1986). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-related measures. Diagnostique, 11, 77-90.

Marston, D., Lowry, L., Deno, S. L., & Mirkin, P. K. (1981). An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study (Research Report No. 49). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

National Council for Accreditation of Teacher Education (2001). A decade of growth 1991-2002. Washington, DC: Author.

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). Antisocial boys. Eugene, OR: Castalia Publishing.

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

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

Shinn, M. R. (1998). Advanced applications of curriculum-based measurement. New York: Guilford Press.

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

Stoiber, K. C., & Kratochwill, T. R. (2000). Empirically supported interventions and school psychology: Rationale and methodological issues, Part I. School Psychology Quarterly, 15, 75-105.

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 Behavior Analysis, 27, 101-113.

Ysseldyke, J. Dawson, S., Lehr, C., Reschly, D., Reynolds, M., & Telzrow, C. (1997). School psychology: A blueprint for training and practice II. Bethesda, MD: National Association of School Psychologists.