Curriculum-Based Measurement- Complete Work

Jane Close Conoley
University of California - Riverside

Jack J. Kramer
University of Nebraska-Lincoln

Conoley, Jane Close and Kramer, Jack J., "Curriculum-Based Measurement- Complete Work" (1993). *Curriculum-Based Measurement*. 11.
https://digitalcommons.unl.edu/buroscurriculum/11

Curriculum-Based Measurement

Buros-Nebraska Symposium
on
Measurement & Testing

Series Editor

JANE CLOSE CONOLEY

**Buros Institute of Mental Measurements
and
Department of Educational Psychology
University of Nebraska-Lincoln**

Curriculum-Based Measurement

Edited by

JACK J. KRAMER
University of Nebraska-Lincoln

Contents

Preface

This volume in the Buros-Nebraska Series on Measurement and Testing provides current information on the development and implementation of curriculum-based measurement. As the title of the volume suggests, effective measurement of children's classroom achievement is not a new problem. Curriculum-based measurement provides an interesting and useful alternative to traditional strategies for assessing academic performance.

This volume continues the tradition of including papers given at the annual Buros-Nebraska Symposium on Testing and Measurement as well as additional contributions selected especially for this book. Each of our authors has made significant contributions to the research that has been produced in the area of curriculum-based measurements.

Stan Deno provides an overview and analysis of curriculum-based measurement (CBM) in the introductory chapter. His paper was presented as the keynote address at the Buros-Nebraska symposium and provides basic information about the manner in which CBM procedures were developed and initially applied to school-based problems. Gerald Tindal analyzes CBM procedures according to nine criteria that have been used to evaluate measurement strategies. This chapter provides a thorough analysis of the technical properties of CBM procedures and a comparison of how CBM relates to other measurement procedures in regard to technical criteria.

Lynn Fuchs demonstrates how CBM can be used to both monitor academic progress *and* improve instructional programs. Her chapter examines the role of CBM within the larger, more complex instructional environment and she suggests specific applications for practitioners and consultants to consider. Ed Lentz and Jack Kramer look at CBM from the perspective of a behavioral model of assessment. A discussion of the basic tenets of the behavioral model is provided and suggestions for future research in curriculum-based approaches is examined.

Ed Shapiro provides a thorough analysis of the implications of CBM for psychoeducational practice. He makes clear his point of view on the value of the entire range of potential applications of curriculum-based assessment (CBA) procedures, "Use CBA." Finally, Mark Shinn and Roland Good conclude the volume with an assessment of the prognosis for the future of CBM. Their chapter provides a fitting summary of the potential benefits of the CBM approach.

Taken collectively, the contributors represent an impressive group of scholars. Their efforts have defined in large part, the curriculum-based measurement approach. The Buros Institute of Mental Measurements is grateful for their time, efforts, and perseverance in completion of this book.

Jack J. Kramer

Curriculum-Based Measurement

Stanley L. Deno

University of Minnesota

In the history of Greek mythology there is a character named Sisyphus who, for sins committed during his lifetime, is condemned to spend eternity pushing a boulder up a hill. No matter how hard Sisyphus tries as he nears the top of the hill, the boulder rolls back down. Sisyphus cannot escape from this continued cycle of effort and failure. Sometimes, when I think about the experiences of many children attempting to learn basic skills in the public schools, I think of the myth of Sisyphus. Too often, it seems to me, no matter how hard they try, they do not succeed.

In this chapter on curriculum-based measurement (CBM) I want to focus on three points: First, what is curriculum-based measurement? Second, why was curriculum-based measurement developed? And third, how does the use of curriculum-based measurement help to avoid the problem of Sisyphus and education?

CURRICULUM-BASED MEASUREMENT DEFINED

As a Subset of Curriculum-Based Assessment

The term *curriculum-based assessment* (CBA) is a very popular topic in the field of special education these days (Tucker, 1985). As Tucker points out, CBA is a term used to describe a practice that has existed for

a very long time—the practice of using what is to be learned as the basis for assessing what has been learned. Since traditional psychometric test construction also involves use of the table of specifications to define the content domains for which test items must be developed, the difference between CBA and traditional psychometric testing may not be immediately obvious. However, four salient differences between CBA and traditional psychometric testing can be identified: 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 are 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 reflects the real goals of instruction in the classroom than most standardized achievement tests, because the assessment information obtained through CBA relates more directly to what is being taught, and also because the content and materials of daily instruction are a fairer and firmer basis for making judgments about student learning.

Since the focus here is on CBM, some clarification of the term is needed. 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 refers to all sorts of information-gathering practices that may occur when observing student performance in the curriculum. These practices include scoring the student's worksheets to obtain a percentage score for the problems or answers correctly completed on a worksheet; making judgments about a student's reading comprehension based on the prosodic features of that student's oral reading; and moving the student to a new skill based on consecutive days of answering all questions correctly. 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 (cf. Howell & Morehead, 1987; Bigge, 1988; Idol, Nevin, & Paolucci-Whitcomb, 1986; Shinn, 1989).

As Distinct from CBA

Curriculum-based measurement (CBM) is a separate and distinct subset of CBA procedures, a specific set of steps for measuring student growth in basic skills, developed at the University of Minnesota through

the Institute for Research on Learning Disabilities (IRLD) (Deno, 1985). The CBM 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 measure frequently and repeatedly the growth of their students in the basic skills of reading, spelling, and written expression. When these procedures are used within the context of the local school's curriculum, they become CBM.

Like CBA, in general, CBM focuses on using existing curriculum materials and goals as a basis for selecting and creating the tasks on which student performance is measured. The primary difference is that CBM is more limited with respect to quantification procedures and types of information collected than is the case with CBA. The term *measurement* in CBM is used to denote the focus on the use of standardization to produce a technically adequate quantitative scale— an issue of less concern in most other CBA models. Although differing in some respects, all curriculum-based approaches share the assumption that data collected from observations of day-to-day student performance in the curriculum are, at the very least, an important supplement for making a broad range of educational decisions. Indeed, a basic assumption is that curriculum-based approaches may be a necessary alternative to commercially distributed achievement tests if measurement is ever going to contribute to educational improvements. Also, curriculum-based advocates generally share the view that traditional approaches to assessment and measurement have failed to contribute sufficiently to educational improvement and that alternatives, such as curriculum-based approaches, offer greater promise.

An Example of CBM

The set of measurement procedures referred to here as CBM were developed through the University of Minnesota IRLD during the years of 1977—83. These research and development activities focused on creating measurement procedures for clearly and simply describing growth in functional literacy. Subsequently, school districts have used similar approaches to develop measures of basic numeracy. Since the focus of all of these research and development activities has been on students who were having significant difficulty developing literacy and numeracy, most (but not all) of the work has been with students in elementary and middle schools. In Figure 1, an illustration of the results

of using CBM procedures with a student in reading over the course of a school year is displayed. As can be seen, student performance in terms of the number of words read aloud correctly in 1 minute from the student's grade level basal reader is presented simply and clearly in relation to changes made in that student's instruction. Although the graph is interesting, the question that needs to be addressed is why so much time and energy were spent to produce such a graph.

WHY CBM?

A Brief Personal History

In the early 1970s at the University of Minnesota, we were attempting to develop a field practicum site that the Special Education Program could use for training resource teachers to serve effectively students classified as mildly handicapped (Deno & Gross, 1973). My role was not only to develop the setting but also to act in the role of practicum supervisor, so I spent my days in a local elementary school working with the students and helping them to develop their intervention skills. An initial problem with which we were faced was how to decide what kind of intervention into a student's program was most appropriate. Although I had my own biases regarding the techniques students ought to use when they were attempting to improve a student's basic skills in an area like reading, I soon discovered that the practicum students had been imbued with a variety of different ideas from different faculty members in their dydactic coursework at the university. I wanted to take a dogmatic position that I as their practicum supervisor had the right to dictate the intervention procedures that they might use; unfortunately, as a scientist, I felt an obligation to remain open-minded regarding the alternatives proposed by my colleagues. After a period of uncertainty regarding how I should approach this task, I decided that the reasonable alternative was to address the problem empirically. The strategy I chose was to allow them to select any of a variety of alternative hunches that they might have regarding how a student might be taught, but to require that students evaluate the effects of whatever hunch they decided to try.

The problem with an open treatment and evaluation approach to making intervention decisions was, and is, how does one evaluate intervention effects with individual students? When teachers evaluate student growth at all, they typically do it on a posttest-only basis. Occasionally, in fields like special education, some effort is made to evaluate intervention effects by doing single-case pre/post

1. CURRICULUM-BASED MEASUREMENT

FIGURE 1
GRAPHIC DISPLAY OF A STUDENT'S READING SCORES
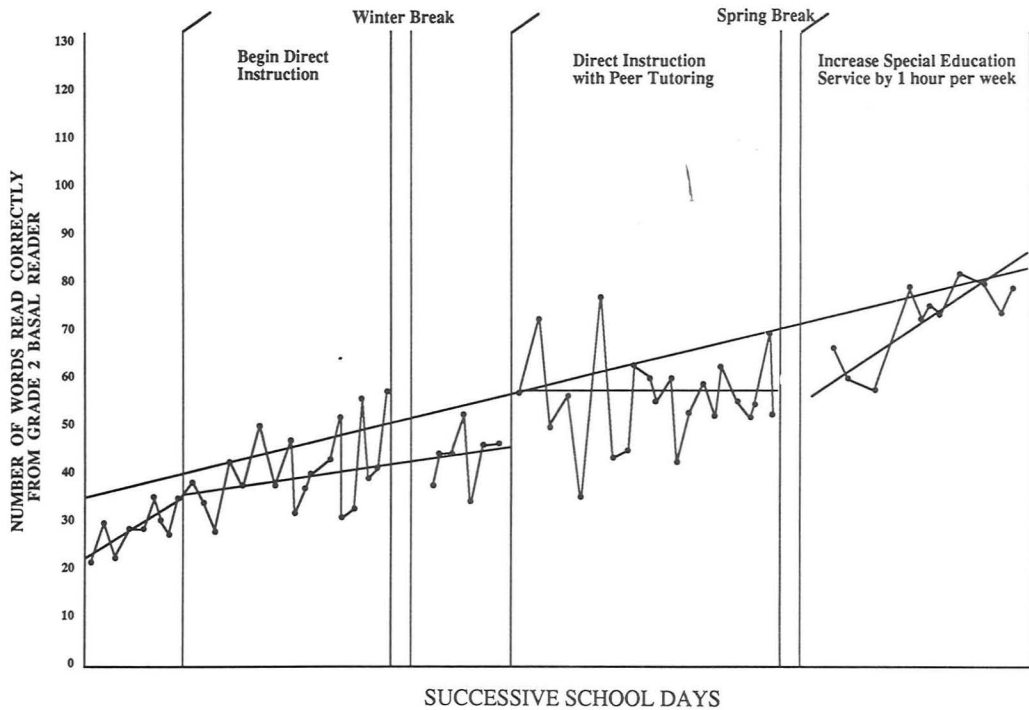
comparisions. As can be seen in Figure 2, however, even though growth may occur during the second phase (as shown by the straight ascending line between pre- and posttesting), our interpretation of that growth will differ, depending upon our knowledge of a child's growth rate prior to the intervention. For Child A, the pre-to-post growth rate is the same as that occurring prior to intervention. For Child B, the pre-to-post growth rate is actually lower than that which occurred prior to intervention. Only in the cases of Child C and D do we have evidence that the students' rate of growth increased in relation to intervention into the children's reading program.

Figure 2
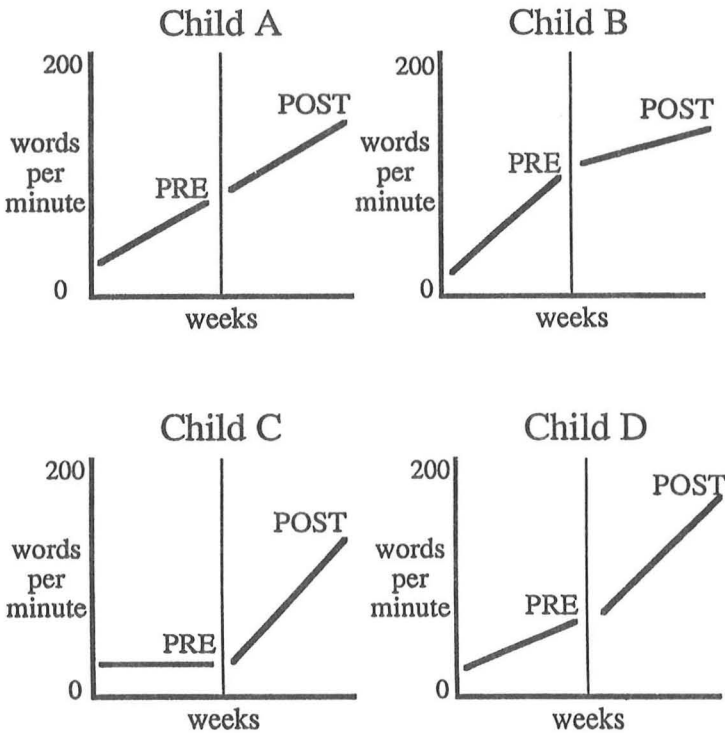
The only feasible solution to the problem of evaluating the effects of interventions with individual students seems to be the use of single-case research design procedures. In single-use research designs, individual performance is measured repeatedly across time to produce a time-series data base that can be used for describing trends in student performance data under different intervention conditions. Thus, in the examples provided in Figure 2, the straight lines representing growth in student performance before and during intervention enable us to make comparative judgments regarding the conditions under which student growth occurs at a higher rate.

A real example of the use of repeated measurement of student performance across time to estimate slope differences in relation to intervention is shown in the data in Figure 3. These data were collected as part of a project to determine whether the effects of special education intervention could be evaluated using the single-case design model (Marston, 1988). As can be seen in Figure 3, both students increase in the rate they are acquiring reading fluency, beginning with the onset of special education. The effects of introducing special education for each student can, in this way, be evaluated, and the general effectiveness of special education can be estimated by aggregating individual cases. The basic schema represented in these two cases, then, provides us with a framework for considering the development of curriculum-based measurement.

Having made the decision to use single-case evaluation procedures to structure special educational interventions, our attention then turned to the development of an ongoing measurement system that teachers could use to establish the kind of data base necessary to produce the evaluation design presented in Figure 3. Since single-case designs require frequent repeated measurement, the question became both what to measure, and how to measure, student performance repeatedly to create the time-series data base required for single-case analysis.

Our initial efforts to develop measurement systems centered upon two approaches. The first approach was a rate of progress measure that was derived from data produced through monitoring the mastery of successive objectives in a sequence of skills or tasks across time (Deno & Mirkin, 1977). Mastery monitoring depends on criterion-referenced measurement of performance on specific tasks or skills typically laid out in a linear or hierarchical order. When using a rate of progress measurement system, the basic datum for evaluating intervention effects is change in the rate at which individual skills are mastered before and after intervention.
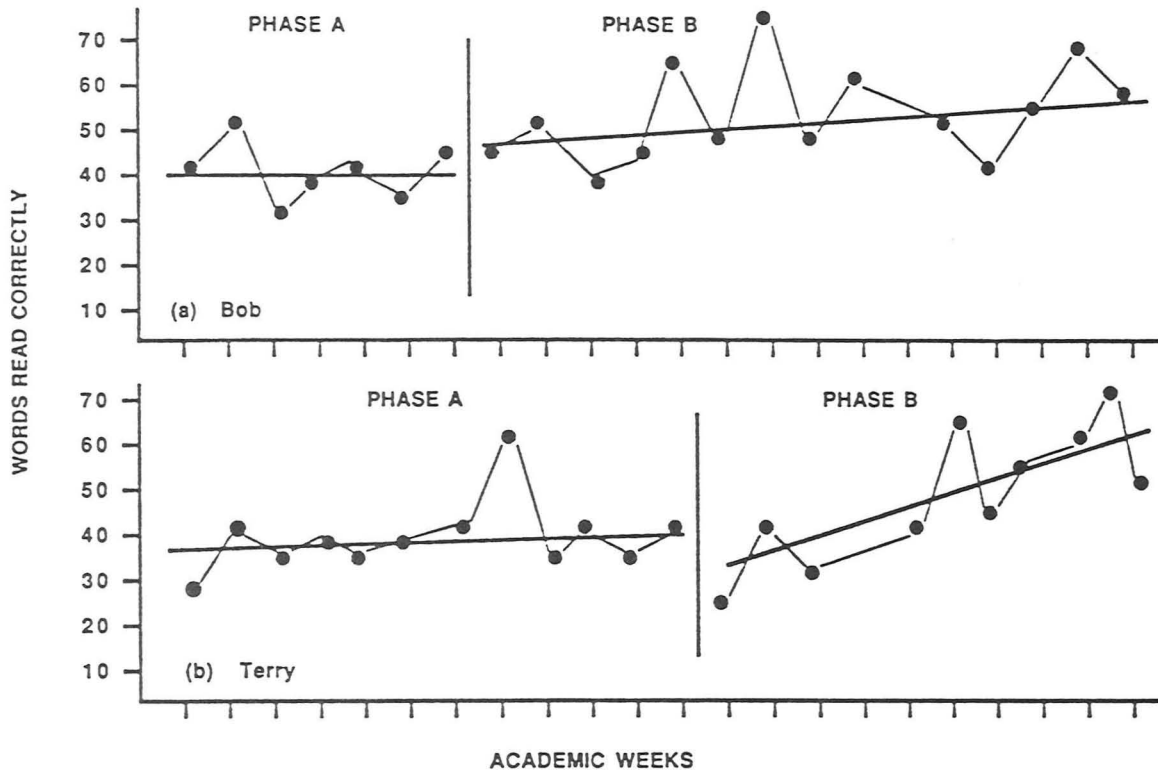
Figure 3. Multiple baseline performance of Bob and Terry

The second approach that we focused on was change in rate of performance on a single task, rather than rate of acquisition, or mastery, of multiple tasks (Deno & Mirkin, 1977). In contrast to the criterion-referenced mastery monitoring approach, this second approach involves specification of a single task on which repeated measurements can be obtained across a very long time period to describe change in proficiency on that task. A good example of measuring performance on a single task is the measurement of the amount of time taken to run a fixed distance, such as one mile. It is common for people who are interested in improving their endurance to monitor closely the amount of time that it takes them to run this fixed distance, and to use changes in the time taken to run the mile as a basis for making decisions about their training program. An analogous measurement system in education might be the length of time that it takes a very young child to print the letters of the alphabet. As a result of our research (Deno, 1985; Deno & Fuchs, 1987), we have come to favor the latter approach—measurement of change on a single task—for purposes of creating curriculum-based measurement procedures.

Reasons for Measuring Change on a Single Task

The rationale for favoring change in performance on an individual task, rather than mastery monitoring across multiple tasks, derives from several disadvantages of mastery monitoring and two advantages for measuring change in performance on a single task.

Mastery as a functional concept. The first problem or disadvantage with measuring the rate of progress in mastering tasks is that the technical and theoretical grounds of the approach are questionable. Three key assumptions must be true for mastery monitoring to be sensible. The first key assumption is that *mastery* as a construct is both theoretically and practically functional in the design and execution of instruction. The issues surrounding this assumption are complex and cannot be adequately considered here. However, the question that must be addressed is whether the acquisition of proficiency in the various curriculum domains actually occurs through mastery of discrete skills; and, following from that, whether instruction should be designed around subskill mastery. If so, then teaching to task mastery and monitoring progress in skill mastery is sensible. However, if student learning can proceed in many different ways for different students (i.e., learning is somewhat idiosyncratic), or if progress in the acquisition of proficiency can occur through partial mastery or skipping of various subskills, then a mastery learning model should not be reasonably imposed upon all students.

A second key assumption that must be met for a progress or mastery monitoring system to be sensible relates to the theoretical question just posed. If all students do not learn, or learn best, by meeting all the mastery criteria in a particular skill sequence, then does it make sense for all students to be required to meet the mastery criterion on each task within a skill sequence before moving on to a new learning experience? The significance of this consideration looms even larger when taking into account the fact that what constitutes mastery on a given task has rarely been empirically established and, therefore, that the mastery criterion specified for each task typically has been stipulated arbitrarily by the curriculum developer. Further, task sequences are almost always logically rather than empirically developed. Thus, the presumed transfer benefits obtained by requiring a student to achieve criterion performance on one task before moving to the next can only be speculations rather than assumptions. When considering these issues, it seems doubtful that teachers should pace their children on this basis. We need to be mindful that theoretical conceptions of children's learning and development ebb and flow, as evidenced by the current return to favoring more "wholistic" approaches. Mastery monitoring, as an assessment approach, is more typically assumed to be aligned with "reductionistic" models that rely on task analysis and isolated skill development. In contrast, CBM procedures function as global indicators of proficiency for different basic skills, and can be successfully used regardless of the particular theoretical conception of learning and cognition underlying curriculum and instructional design. For us, this has meant moving away from mastery monitoring systems that must be wedded to a particular approach to curriculum and instructional design.

A third key assumption that must be met for mastery monitoring systems to be sensible is that they be both technically and logistically feasible within the context of everyday instruction in the schools. The advent of microcomputers in the schools has made it possible to manage relatively complicated data sets in the classroom that can provide teachers with information on individual student progress. At present, however, the amount of information that teachers must process when monitoring individual student performance across several different basic skills exceeds practical limits. Further, as the number of subskills on which students are measured increases, the logistical problems increase for the teacher. Given advances in technology, this problem is not insurmountable; however, with CBM procedures we have tried to develop an approach that can be used in the current classroom without waiting for technological development.

Fractionation of learning. A second important problem associated with the mastery monitoring approach is that it fractionates the essential outcomes of learning in a particular curriculum domain. Thus, for the student, and often for the teacher, reading becomes performance on a series of isolated tasks represented as questions and answers on worksheets and curriculum-embedded mastery tests.

Too often, I am afraid, the result of this focus on isolated elements of the curriculum produces confusion in the minds of both teachers and students over the essential nature of what is being learned. Indeed, this overemphasis on the details of daily lessons is very likely what led Charles Silberman (1970), in an earlier call for educational reform, to identify "mindlessness" of educators regarding the purpose of education as the central problem of the schools. Students are affected, as well, by this fractionation of the curriculum. The dialogue between a special education teacher trainee, Diane, and her son, Ben, that is presented in Figure 4 illustrates what is most probably a common student viewpoint. The difference is that Ben is a very perceptive and articulate 7-year-old who seems to have reconciled the discrepancy between what his teacher does in the name of reading at school and what he has learned reading to be at home. In this dialogue, Ben makes clear that what he has learned to enjoy in the name of reading at home has very little to do with what he is required to do in the name of reading in school.

Figure 4

Conversation between Ben (7 years) and Mom, Fall '86

How come you always ask people about what they mean when they ask you if you like reading?
    You know. Reading at home or reading at school.
Aren't they the same?
    No. Like bat and bat.
Bat and bat?
    Yeah. You know. A bat like a thing that flies in caves and a bat you hit a homerun with.
What does that have to do with reading at school or at home?
    It's just the same. Reading and reading. You know.
No. I don't know. When you read at home you look at pages of a book, read the words, and find out from the words what happens. Isn't that the same as any reading? Isn't that the same thing reading is at school?
    No. At school reading is looking at charts and doing worksheets and workbook and book and the teachers talks and stuff. You know. You're sposed to get them all right. Not fun.
Yes, but the book part, isn't that the same as reading other places?
    No. You can't choose the stories and if you like it, it's not fun.
Why not, if you like the story you read.
    Cause you can't finish it. If you do, you'll get in trouble.
Why? You're not supposed to go ahead of others?
    Yeah. But I would get in trouble because if I sneaked and read the end, I wouldn't have time to finish my work.
                     D. Lilleberg

Although we might predict with some certainty that Ben will survive his school experience in reading, we may also speculate that the disinterest in reading exhibited by many secondary students and the shamefully high proportion of illiteracy among American adults has occurred because they became lost very early in the trees of their school's reading curriculum and never experienced the beauty of the forest that we know as reading.

Skill sequences as independent variables. The third problem we encountered when using mastery monitoring approaches was that we could not use the data generated through measuring student progress on the objectives to evaluate the use of alternative skill sequences. This problem occurs because, in mastery monitoring, the rate of progress on the skill sequence functions as the dependent variable. That is, mastery of the skills in the hierarchy defines the outcome, rather than the inputs, of instruction. If, we were interested in using the data generated through mastery monitoring to evaluate the use of a different curriculum that included a very different skill sequence, we could not do so. In effect, when one adopts a particular mastery monitoring system, one also adopts a particular scope and sequence of skills as the essential objectives of instruction. We wanted teachers to have data representing changes in dependent variables, independent from particular curriculum sequences, that could serve in evaluating alternative curricula and sequences of objectives. To do so required measurement procedures that were not wedded to one curriculum sequence.

Clarifying the focus. A fourth reason why we have opted for a measurement system based on measuring change in performance on a single task across time is that repeated measurements on the same task aids in focusing attention on an important proficiency indicator. This point, of course, is related to the "forest and the trees" problem, but the emphasis here is on the need for teachers to have clear and unambiguous feedback regarding the general effects of their instructional efforts. Too often, I think, teachers are either uncertain about the overall effects of their efforts to teach basic skills, or they are certain that they have been successful when a student has mastered the particular skills they have been teaching. In the first instance, their uncertainty stems from the fact that they have no "vital sign" indicators, such as pulse rate and temperature, that they can use to monitor the effects of their treatments on the educational health of their students. Indeed, I sometimes think teachers are like early flyers 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. In the second

instance, teachers' excessive certainty in their success stems from their overconfidence that specific skill mastery can be taken as evidence that the student is increasing in proficiency in the general curriculum where that skill is being taught. Since very little empirical justification ever exists for such an inference, the risk is real that teachers will conclude, as did the misguided surgeon, that "the surgery was a success, but the patient died."

Technical characteristics. Our final reason for building measurement procedures around change in performance on a single task was that the technical characteristics of such a system were superior to those of mastery measurement. Two facts, in particular, led to this conclusion. First, in our early efforts to assist teachers in developing and using progress measurement systems based on mastery measurement, we found the scores to be unreliable. This occurred because teachers tended to be inconsistent in their application of the mastery criteria— often for practical reasons, such as wanting to "keep the student up with his group," but sometimes for altruistic reasons, such as, "He came so close, I didn't want him to feel bad." Although each of us can appreciate why such reasons operate to produce variation from the mastery standard, it does not alter the fact that the data produced are of unknown reliability.

The second fact that led to our conclusion that scores based on repeated measurements of performance on a single task were preferable to those produced through mastery monitoring was that the scores produced by the former method were based on more nearly equivalent behavioral units than those produced when plotting progress in mastery of diverse skills. It is unreasonable to equate two separate reading subskills such as "identifying initial consonant blends" and "reading words with prefixes and suffixes" either behaviorally or cognitively. Any effort to plot graphically the mastery of these two tasks across time will most certainly reveal that students will take longer to master one than the other. When task or skill hierarchies are composed of a heterogeneous mix of skills of differing difficulty, it becomes virtually impossible to rely on a scale showing individual student progress in successively mastering those tasks, and to use graphs of student progress across time for evaluating the effects of changes in a student's instructional program. We believe that the actual performance scores obtained by repeatedly measuring student performance on the same task, using CBM procedures, are technically superior and more directly interpretable.

Selecting Tasks for Performance Measurement

Once we had established that our CBM procedures were to be based on repeatedly measuring performance on the same task, the importance of selecting the tasks for measurement became obvious. To stipulate and measure arbitrarily on tasks of unknown validity—so often the case when informal curriculum-based assessment occurs—would be indefensible.

A two-part strategy was used to identify those tasks that teachers might use in CBM. The first part of the strategy—initial task selection—was based on research using a criterion-validity paradigm to select those tasks that seemed to be the best candidates for repeated performance measurement (Deno, 1985). The second part of the task selection strategy was to test the tasks' instructional utility by evaluating the student achievement of teachers using the CBM data to make instructional evaluation decisions (Fuchs, Deno, & Mirkin, 1984; Fuchs, in press).

Criterion validity. In selecting tasks for the criterion-validity research, practical parameters were established for considering a task as a viable candidate for CBM:

1. *Repeatability.* Since the goal of measurement was to create a graphic time-series record of change in student performance, a task had to be one on which frequent repeated measurement could occur.
2. *Multiple forms.* Since repeated measurement was to occur and change in performance was to represent real growth in general proficiency, rather than the effects of practice on a specific task stimuli, a task had to be one for which it was simple to create many equivalent forms.
3. *Inexpensive.* Since many forms had to be made available for teachers to use frequently, the task had to be one that would not require costly materials.
4. *Time efficient.* Since frequent repeated measurement was required to create the graphic time-series record, the task needed to be one that did not consume too much instructional time.
5. *Easy to teach.* Since many teachers, paraprofessionals, and possibly students were to administer the measures, the task had to have one for which simple measurement procedures could be created and easily taught to nonprofessionals.
6. *Reliability.* Since the data were to be used to make important instructional intervention decisions, the tasks had to be ones for which reliable measures could be constructed.

Establishing parameters in task selection was important in the early program of research and development on CBM because it delimited the range and variety of tasks included in our search for valid indicators of reading proficiency. In addition, specifying the characteristics of a practically feasible task on which to do frequently repeated measurement enabled us to focus our criterion-validity research on only those tasks that could be part of a classroom-based, ongoing formative evaluation system.

The reasons for limiting task selection have not always been fully understood or appreciated by many, however. Indeed, the failure to include tasks for measurement that might operationally define the measurement domain more broadly is often mistakenly used as a basis for asserting that the CBM measures are invalid. A good illustration of the problem is in the area of reading, where we identified "reading aloud from text" as a task that can be used to create a global indicator of reading proficiency (Deno, Mirkin, & 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. The criterion validity research (summarized in Shinn, 1989) on using this task in reading measurement provides a solid empirical basis for concluding that the number of words read aloud correctly from text in a 1-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' ages and grades, to teachers' judgments of reading proficiency, and to teachers' placements of students 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 as well as so-called "direct measures" (cf. Fuchs, Fuchs, & 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 measurement procedures we have been developing. To use either task would (a) consume far too much time to be used in a frequent measurement system (students would have to read fairly lengthy passages so that question asking or story retelling would be sensible); (b) cost too much in the development of multiple equivalent forms; and (c) in the case of story retell, be difficult to teach others to score reliably. Thus, although these tasks have been used as criterion measures in our

validity research, they were excluded as candidates for our CBM procedures on other important grounds. We have painfully learned, however, that neither empirically nor technologically valid reasons are enough to persuade the critics. Clearly, face validity reigns supreme in education. A measure had better meet the consumer's preconceived notions of what an operational definition of the construct is supposed to look like if it is to be accepted easily. One cannot help but wonder if chemical engineers initially resisted the use of litmus paper because the "colors weren't right," or if doctors wouldn't use thermometers because they believed that a patient would "feel warm" if suffering from a fever. At the very least, we must conclude that, when it comes to measurement, educators are radically behavioral—operating as if inference beyond directly observed behavior is inappropriate.

Instructional utility. The criterion-validity data led us to conclude that it would be possible to teach teachers to use the CBM procedures to monitor routinely student performance and to evaluate the effects of daily instruction using the data thereby produced. Our hypothesis was that teachers using frequently collected data that graphically illustrated the rate of change in student performance could become more effective in timing their instructional change decisions, and that the result would be increased student achievement. To test this hypothesis, we designed a comparative study, in which special education teachers who used CBM in formatively evaluating their instruction were compared to teachers who used more conventional procedures (Fuchs, Deno, & Mirkin, 1984). The results of this study confirmed the hypothesis that teachers could increase students' achievement using the CBM procedures in formative evaluation. An important related outcome of the research was that evidence was obtained revealing that increases in CBM scores were related to increases in standardized achievement test scores, and most importantly, that increases in the number of words read aloud correctly in 1 minute across the school year were directly related to increases in the reading comprehension subtest scores of the students.

APPLICATIONS OF CBM

The results of the CBM research program have provided a basis for developing standardized measurement procedures that can be used to evaluate formatively 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 (Fuchs, Deno & Mirkin, 1984; Fuchs, in press). 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. These decisions include screening and monitoring high-risk students in the regular classroom program (Marston, 1988; Espin, Deno, Maruyama, & Cohen, 1989), evaluating prereferral interventions (Marston & Magnusson, 1985; Shinn, 1989), and developing IEPs (Deno, Mirkin, & Wesson, 1984), as well as reintegrating and follow-up monitoring of students terminated from special education services (Allen, 1989).

Advantages of CBM. Because traditional achievement measures have been used to "data-base" educational decisions for years, it is fair to wonder what the advantages of CBM might be. Several can be identified. First, because CBM data can be used to measure frequently performance across relatively short time periods, a new metric—slope— is available to evaluate interventions into individual student programs. The advantage of the slope metric is that it can be used to contrast the rate change in individual student performance under various instructional programs. Thus, teachers can execute a program, examine its effects on the rate of academic growth, change the program, examine the effects of the change relative to the previous program, and then decide whether to continue with the new program or to restore elements of the previous program. The continuous feedback regarding slope at various times enables teachers to make ongoing, data-based instructional decisions that are responsive to individual students. The net effect of using the slope data in this manner should be to improve cumulatively individual student programs. An illustration of the use of CBM data to improve cumulatively a student's program is presented in Figure 5. This figure is a graphic portrayal of the number of words read aloud correctly in 1 minute by Candy from his grade basal reader. Each heavy vertical line drawn on the graph identifies the point where a deliberate change was made by his teacher in an effort to find a more effective means of teaching him to read. The straight lines drawn through the data between vertical lines are a visual representation of the slope of Candy's performance during that phase of his program. As is evident from an overall inspection of Candy's progress, some of the changes introduced by his teacher into his program are associated with increases in slope and some are associated with decreases. Toward the end of the year, however, the overall trend in Candy's performance is increasing more rapidly than it was during the first half of the year. We cannot be certain that this more rapid rate of increase in performance is the result of his teacher's use of CBM data to continually evaluate his program

and modify it in response to his performance; nevertheless, this is a plausible inference consistent with the research on the increased instructional effectiveness of teachers using CBM data in formative evaluation.

NUMBER OF WORDS READ ALOUD

Figure 5

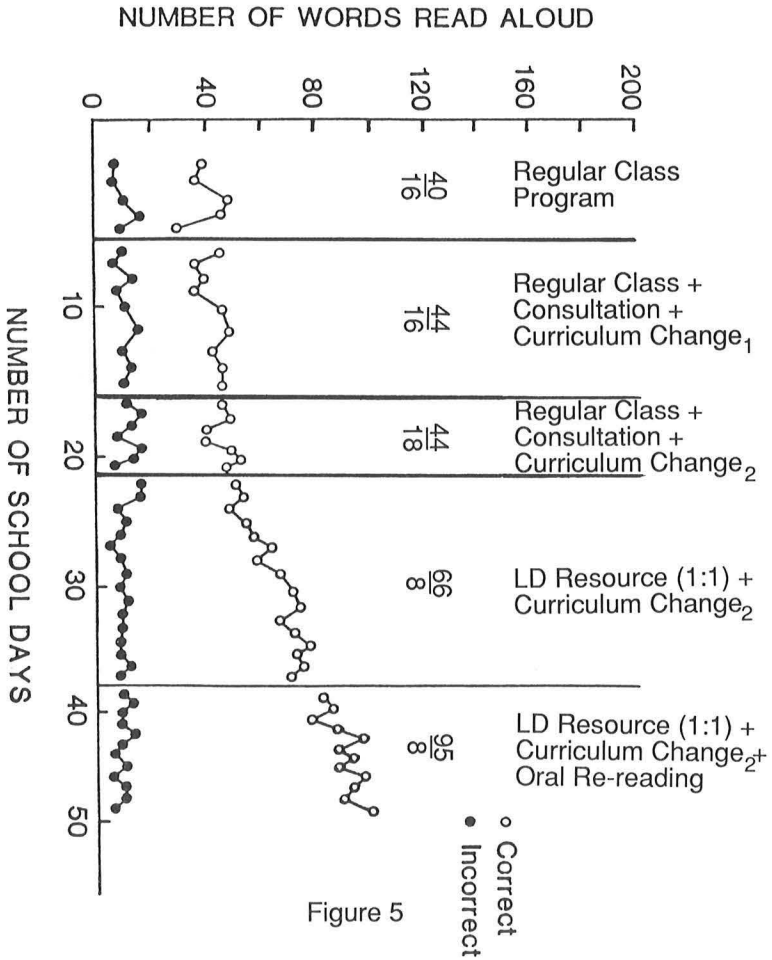A second advantage of the CBM data is that they can more easily be used to communicate an individual student's progress in reading than is typically the case with commercially available standardized tests. This ease in communication derives from both the nature of the data presentation in CBM and the additional references available when CBM is set in the larger context of an ongoing evaluation system. The

clarity of data presentation and interpretation is evident in viewing Figure 5. The number of words read correctly and incorrectly in 1 minute of reading from standard classroom text is not a datum that requires much explanation. Further, the simple line graph showing calendar dates and weekdays clearly reveals the level, trend, and variability of performance in student performance relative to significant periods of the school year. The utility of these graphs in communication was illustrated in the data collected by Fuchs, Deno, and Mirkin (1984). In that study comparing the effects of teachers using CBM data in systematic formative evaluation of individual students' programs, both the teachers and the students were able to specify correctly not only the students' IEP goals in reading, but also were able to predict accurately whether or not the students were going to make their goals. Comparison teachers using more conventional approaches to writing IEP goals and evaluating students' progress toward those goals could neither specify the goals at year's end, nor could they and their students correctly predict whether those goals would be attained. A strong argument can be made that a data system needs to be well and easily understood by those who are using it, if it is to become a functional part of students' programs.

CBM data graphs also communicate clearly because of the increased meaningfulness resulting from the increased number of references available when examining a student's graph. First, a student's performance is *curriculum referenced* in that the data reveal level, change, and variability in student performance on standard text material drawn from the student's local school and classroom. Second, a student's performance is *goal (or criterion) referenced* in that day-to-day performance can be compared both to the goal specified on the graph and to the daily increase required to attain that goal on the date specified for goal attainment. Third, a student's performance is *individually referenced* in that we can easily contrast the level, trend, and variability of the student's current performance with that same student's past performance. Fourth, student performance is *program referenced* in that it reveals how well the student progressed under different program arrangements or methods. Finally, a student's performance can be *norm referenced* by displaying how well a representative sample of that student's peers are doing in reading from the same material at the same time. A reading of Candy's graph in Figure 5 reveals all five types of references available in an individual student's CBM data graph. This rich array of referencing, easily and quickly apprehended in the graphic display of Candy's CBM data, becomes a powerful tool in the important communications surrounding an individual student's success in school.

Problems in implementing CBM. To describe CBM as if it is a measurement alternative with no associated problems or disadvantages would be misleading. In an effort to identify clearly the major barriers to implementing CBM, we conducted a Delphi survey of administrators and teachers who had implemented and were using CBM in their administrative units. The results of their inquiry are presented in Figures 6 and 7. A comparison of the administrators' and teachers' responses reveals a number of interesting differences. Teachers focus on the immediate impact of using CBM on a frequent basis and express concern about the additional time required in doing CBM. Three of the five most frequently identified barriers by teachers refer to time-associated problems. The remaining two teacher concerns relate to issues of measurement validity. As mentioned previously, the criterion-validity research rarely is powerfully persuasive with the teachers, and the face validity of CBM in reading and written expression is not high enough for many teachers. At the same time, less than 15% of the teachers who responded in the survey said they thought it was not a good idea that their district had implemented CBM.

The administrators' view of problems associated with implementing CBM was quite different from that of the teachers. The emphasis in the administrators' 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 teachers' 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 administrators' perspective was the natural resistance that occurred when any change in practice was required of school personnel.

CONCLUSION

Curriculum-based measurement (CBM) has been presented here as an alternative to the more conventional measurement approaches available to educators—particularly special educators. Like curriculum-based assessment (CBA), CBM relies on direct observation of student performance on stimulus materials drawn directly from the local school curriculum. CBM is distinct from CBA in its specification of both what should be measured (i.e., the tasks) and how measurement should occur (i.e., the procedures). The gains accruing through the standardization used in CBM are those typical of improved technical adequacy in measurement: increased reliability and validity of the information obtained through measurement. Further, standardization
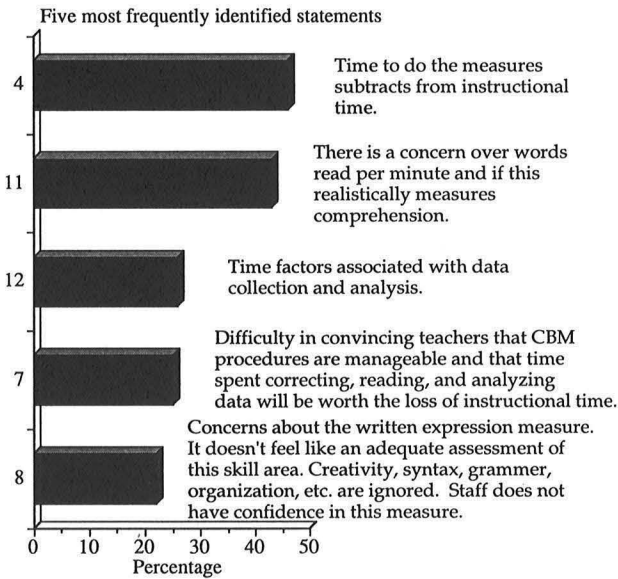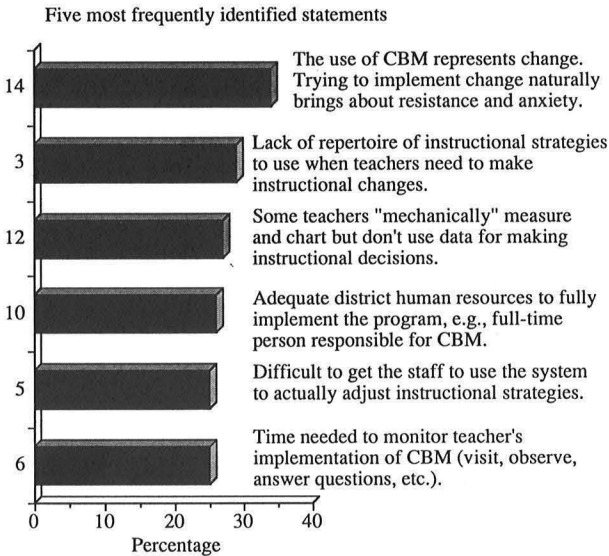
Figure 6

Delphi Probe - Teachers

Five most frequently identified statements

4	Time to do the measures subtracts from instructional time.
11	There is a concern over words read per minute and if this realistically measures comprehension.
12	Time factors associated with data collection and analysis.
7	Difficulty in convincing teachers that CBM procedures are manageable and that time spent correcting, reading, and analyzing data will be worth the loss of instructional time.
8	Concerns about the written expression measure. It doesn't feel like an adequate assessment of this skill area. Creativity, syntax, grammer, organization, etc. are ignored. Staff does not have confidence in this measure.

0   10   20   30   40   50
Percentage

Figure 7

Delphi Probe - Administrators

Five most frequently identified statements

14	The use of CBM represents change. Trying to implement change naturally brings about resistance and anxiety.
3	Lack of repertoire of instructional strategies to use when teachers need to make instructional changes.
12	Some teachers "mechanically" measure and chart but don't use data for making instructional decisions.
10	Adequate district human resources to fully implement the program, e.g., full-time person responsible for CBM.
5	Difficult to get the staff to use the system to actually adjust instructional strategies.
6	Time needed to monitor teacher's implementation of CBM (visit, observe, answer questions, etc.).

0   10   20   30   40
Percentage

permits aggregation of data across students for general program evaluation and establishes the conditions necessary for norm referencing. No gain is without loss, however. Standardization and prescription in measurement reduce the flexibility available through direct observation and recording of student behavior in the classroom and curriculum. The obvious solution to the problem of potential loss when using CBM is, of course, to train educators to use CBM and whatever other measurement procedures are appropriate in each individual case.

CBM has been developed to provide teachers with the tools to evaluate formatively the instruction they are providing to students who are developing functional literacy and numeracy. The goal has been to design procedures that teachers could use to make informed instructional decisions in such a way that they effect higher levels of achievement in their students than would otherwise be the case. Research evidence has accumulated that achievement increases can occur when teachers use CBM procedures to "data-base" their instruction. The research also makes clear that the connection between the simple collection of CBM data and increased achievement is not direct and automatic. The teacher's competence in using the data and designing alternative instruction mediates this relationship. When the CBM data signal the need for program change to a resourceful teacher, that teacher introduces program modifications that increase student success. The same signal sent to teachers who either are constrained by circumstances making change in students' programs impossible, or to teachers who do not know what else to do when a student is not learning, will not result in increased student achievement. There is no escaping from the fact that competent people are only made better when they use improved tools for doing their work and have the time and resources required for success.

REFERENCES

Allen, D. (1989). Periodic and annual reviews and decision to terminate special education services. In M. Shinn (Ed.), *Curriculum-based measurement*. New York: Guilford.

Bigge, J. (1988). *Curriculum-based instruction*. Mountain View, CA: Mayfield.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219-232.

Deno, S., & Fuchs, L. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*, 1-16.

Deno, S., & Gross, J. (1973). The Seward-University Project: A cooperative effort to improve school services and university training. In E. Deno (Ed.), *Instructional alternatives for exceptional children*. Reston, VA: Council for Exceptional Children.

Deno, S., & Mirkin, P. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.

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

Deno, S., Mirkin, P., & Wesson, C. (1984). How to write effective data-based IEPs. *Teaching Exceptional Children, 16*, 99-104.

Espin, C., Deno, S., Maruyama, G., & Cohen, C. (1989). *The Basic Academic Skills Samples (BASS): An instrument for the screening and identification of children at risk for failure in regular education classrooms*. Unpublished manuscript available from S. Deno, Burton Hall, 178 Pillsbury Dr. S.E., Minneapolis, MN 55455.

Fuchs, L. (in press). Enhancing instructional programming and student achievement with curriculum-based measurement. In J. Kramer (Ed.), *Curriculum-based assessment: Examining old problems, evaluating new solutions*. Hillsdale, NJ: Erlbaum.

Fuchs, L., Deno, S., & Mirkin, P. (1984). Effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*, 449-460.

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

Howell, K., & Morehead, M. K. (1987). *Curriculum-based evaluation for special and remedial education*. Columbus, OH: Merrill.

Idol, L., Nevin, A., & Paolucci-Whitcomb, P. (1986). *Models of curriculum-based assessment*. Rockville, MD: Aspen.

Marston, D. (1988). The effectiveness of special education: A time series analysis of reading performance in regular and special education settings. *Journal of Special Education, 21*(4), 13-26.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52*(3), 266-276.

Shinn, M. (1989). *Curriculum-based measurement*. New York: Guilford.

Silberman, C. (1970). *Crisis in the classroom*. New York: Random House.

Tucker, J. (1985). Curriculum-based assessment: An introduction. *Exceptional Children, 52*(3), 199-204.

A Review of Curriculum-Based Procedures on Nine Assessment Components

Gerald Tindal

University of Oregon

The purpose of this chapter is to describe curriculum-based procedures from a broad perspective that encompasses the major models appearing in the professional literature in the past 10 years. Rather than simply review the major perspectives, operating assumptions, and implementation directives of these models, however, nine criteria are presented for a uniform comparison. These criteria were implicit in the adoption of curriculum-based measurement (CBM) in Pine County, Minnesota during the initial training and field-based research conducted in the early 1980s. Therefore, they can be used both to structure the review and to provide district personnel a focused evaluation strategy for adopting any or all components of the models.

MODELS OF CURRICULUM-BASED PROCEDURES

Curriculum-based assessment (CBA) has been variously defined in the professional literature since it was originally introduced in the early 1980s. Although many of these definitions include similar components, the differences between them are sufficient to warrant a careful examination. In part, the models can be compared by analysis of their conceptual base and assumptions, the essential features that comprise

any specific model. However, a more important comparision of the models may be in the empirical and psychometric support that exists.

An immediate problem that must be resolved is agreement on a definition of curriculum-based procedures that is generic enough to encompass the various models. The key words in the phrase are "curriculum-based" and "assessment." I will confine the first term to the analysis of the materials used for measuring and evaluating student performance and the second term to the collection of information for making a decision. This second term, therefore, refers to several issues: the type of response that students make, the system for scoring and reporting performance, and the interpretations or type of decisions that can be made from the data. The only other criterion for considering a model of curriculum-based procedures is that it must be presented or described in the professional literature (with or without supporting data). With this basic definition, several major models of curriculum-based assessment can be considered. Although not all models explicitly employ the term curriculum-based assessment, they in fact represent measures of student performance that fit the basic definition above.

Gickling and Havertape (1981), Gickling and Thompson (1985), Tucker (1985), and Coulter (1985) have all written about a consistent model of curriculum-based assessment. This model is more explicitly developed in reading and mathematics, but has been extended to other areas. It is very closely linked to instructional planning for individual students with three major dimensions: (a) task type ("context" tasks and drill tasks), (b) task items (knowns, hesitants, and unknowns), and (c) performance levels (frustration, independent, and instructional).

Idol, Nevin, and Paolucci-Whitcomb (1986), Blankenship (1985), and Bursuck and Lessen (1985, 1987) have described a model that is very closely connected to criterion-referenced testing (Popham, 1972, 1984; Berk, 1984) and spans a wide range of basic skills and content areas. The procedures used to create measures of learning in specific domains (defining the domain, selecting an item sampling strategy, and establishing criteria of success) are considered in this model. Howell and Morehead (1987) also describe a model of curriculum-based evaluation (CBE) that is similarly organized with domain referencing and criteria for mastery, though they focus more on basic skill areas and less on content knowledge.

Precision Teaching (Lindsley, 1964; White & Haring, 1980) provides a model of direct assessment using curriculum-based procedures that has been in operation for over two decades. This model uses task-analyzed skill sequences and a standard behavior chart to evaluate instructional programs.

Finally, curriculum-based measurement (CBM) has appeared from the work of Deno and Mirkin (1977) and was expanded through the research conducted at the Institute for Research on Learning Disabilities (Deno, 1985, 1989; Germann & Tindal, 1985; Fuchs, 1989; Marston & Magnusson, 1985; Shinn, 1989; Tindal, 1989). The work of Shapiro (e.g., 1989) is consistent with many of the operating procedures of CBM.

Although the distinctions noted above imply that the various models are quite different, there is, in fact, a considerable amount of blurring and cross-fertilization across models. For example, the work of Bursuck and Lessen (1987) and Rosenfield (1987) has obvious components of CBM mastery monitoring (Deno & Mirkin, 1977); Howell and Morehead (1987) have elements of this early work and precede some of the later work of CBM (Shinn, 1989). The original work of Deno and Mirkin (1977) has components that Idol et al. (1986) have developed more fully. So, the distinctions that are made in comparing the different models should not be taken as black and white, but rather as shades of gray. These distinctions, nevertheless, are important and have implications for use by individual schools or school districts.

COMPONENTS OF CURRICULUM-BASED PROCEDURES

Nine components of assessment are used to compare the models, selected to both accommodate the various models and to reflect relevant features known to influence decision-making in the schools. Following is a brief listing and description of each component.

1. *Focus of behavior within the assessment process.* Two dimensions of student behavior are considered, basic skills or a content knowledge focus.
2. *Curriculum-based item sampling.* Since all models employ the term curriculum-based, it is imperative that some definition be given to both the curriculum and the manner in which items are sampled for inclusion in assessment devices.
3. *Administration and scoring procedures.* An important component in all measurement is the manner in which assessment devices and instruments are implemented.
4. *Type of response.* This component is closely related to administration and scoring (which focus on the stimulus materials), with two responses considered: production and selection.
5. *Technical adequacy.* All assessment and measurement must conform to the standards established by the American Psychological Association, American Educational Research Association, and National Council on Measurement in Education

(1985). Thus, this criterion is included as a component for reviewing curriculum-based procedures.

6. *Frequency of measurement.* Implied in all assessment and measurement activities is a schedule or frequency of data collection. This component has bearing, in turn, on the manner in which student performance data are summarized and how data are used.

7. *Display of data.* The manner in which data are displayed has important bearing on how they are used; this component is implicit in most of the models of curriculum-based procedures.

8. *Reference guides for data interpretation.* All numbers must be anchored to some type of reference or comparison, in order to provide a meaningful interpretation. Three specific references are considered: (a) norms, (b) criterion (absolute standards), and (c) previous performance.

9. *Use in decision making.* Eventually, all curriculum-based procedures are used to help educators make decisions; however, the decisions for which they are applicable differ, in great part because of the previous components.

These nine components form the backbone of the following review. The different models are analyzed according to their consideration of each component, both implicitly and explicitly. Some curriculum-based models, although not espousing one of these components as a major tenet, provide a strong commitment to it nevertheless.

FOCUS OF BEHAVIOR WITHIN THE ASSESSMENT PROCESS

All assessment can focus on either (a) skill mastery or (b) content knowledge. These two terms should not be considered as categorically distinct, but at polar ends on a continuum, as depicted in Figure 1 below.

Skills are defined with motoric responding as the essential feature. At their extreme, they may be considered tool movements (White & Haring, 1980), which are physical behaviors necessary for functional application of more advanced behaviors. For example, speech sounds and blending are tool movements for oral reading; pencil holding/ movement and number formation are tool movements for math computation solving. The other dimension of skills is the inclusion of both accuracy and rate as important dimensions that together comprise automaticity, or fluent responding in the presence of distractors (Howell & Morehead, 1987).

In contrast, "knowledge originates in information which can be received directly from observations or indirectly from reports of observations. These observations may be external (objects or events) or internal (thoughts and feelings)" (Sheffler, 1965, p. 137, as cited in Ebel, 1982). As Ebel continues to expound, information must be manipulated to become functional knowledge, so two further distinctions are proposed: (a) the type of expressions in which information is conveyed—facts, concepts, or principles (Roid & Haladyna, 1982)—and (b) the format in which information is expressed, using oral or written communication systems (Tindal & Parker, 1989).

It is generally assumed that learning moves from an emphasis on skills to knowledge and manipulation of information. In the early elementary school years, students learn basic skills of math computation, reading, spelling, and written communication; later, in the intermediate years (middle and high schools), this emphasis on basic skills is replaced with a focus on information in content areas, such as geography, earth science, algebra, geometry, etc.

The different models of CBA, CBE, and CBM differ considerably in the attention devoted to either basic skills or content knowledge. The various authors, however, have not really addressed such a distinction directly, so the following statements represent assertions derived from the professional literature.

On the skills end of the continuum are advocates of curriculum-based assessment (Gickling & Havertape, 1981), precision teaching, curriculum-based evaluation (Howell & Morehead, 1987), and curriculum-based measurement (Deno, 1985, 1989). For example, measurement probes described by White and Haring (1980) include students printing letters or numbers as fluently as possible (accurately and correctly). The measurement system described by Gickling and Thompson (1985) includes student oral reading and placement into levels which parallel those of an informal reading inventory (frustration, instructional, and independent). The research conducted on CBM has generally focused on well-defined behaviors that are generally on the skills end of the continuum. In fact, the initial research that began this line of investigation focused on the development of measures that were (a) technically adequate, (b) capable of frequent administration, (c) easy to learn to administer and to teach others to administer, and (d) capable of generating many parallel forms (Deno, Mirkin, & Shinn, 1979). These criteria were considered in developing a broad measurement net in the basic skill areas during the initial studies (Deno, Mirkin, Chiang, & Lowry, 1980; Deno, Mirkin, Lowry, & Kuehnle, 1980; Deno, Mirkin, & Marston, 1980). The data from these studies supported the following

behaviors as reliable and valid indices of student performance: (a) in reading, the number of words read correctly; (b) in written expression, the number of words written or words spelled correctly; and (c) in spelling, the number of words spelled correctly and the number of correctly sequenced letters. In the basic skill areas, assessment generally encompasses more diverse behavior samples than those represented in CBM; furthermore, content areas are included within the assessment focus. For example, Idol et al. (1986) describe construction of questions to be asked following a reading sample similar to that used with the Informal Reading Inventory (IRI) (Johnson, Kress, & Pikulski, 1987). However, recent research has focused on written retell of passages (Fuchs, Fuchs, & Maxwell, 1988; Tindal & Parker, 1989).

On the content knowledge end are most other advocates of CBA (Idol, Nevin, & Paolucci-Whitcomb, 1986) and criterion-referenced testing (Berk, 1984). Such measurement systems address a number of issues such as defining the domain, sampling items from that domain, and determining mastery within the domain. These authors delineate procedures for constructing tests in more content-specific areas, such as science and subareas in mathematics. Howell and Morehead (1987) and Tindal and Marston (1990) describe a number of procedures for assessing reading comprehension, including maze, cloze, and retellings.

This dimension is portrayed in Figure 1. On one end of the continuum is a skill focus and on the other end is a knowledge or information focus. At the bottom are descriptors of general features of each end and an example of their extremes. Clearly, any content can be considered from either end of the continuum. Instruction and assessment can focus on teaching and learning rules and factual information by employing them in actual communication systems (i.e., spelling words correctly and efficiently while writing) or reiterating them as static information (i.e., the rule for doubling consonants when adding suffixes).

CURRICULUM-BASED ITEM SAMPLING

Although all models of curriculum-based procedures imply that measurement items are derived from the curriculum, a wide variety of sampling plans are nevertheless available.

Most advocates of curriculum-based assessment treat the curriculum for instruction and that for assessment as isomorphic. For example, Tucker (1985) states that "curriculum-based assessment is the ultimate in 'teaching the test,' because the materials used to assess progress are always drawn directly from the course of study" (p. 200). The item-sampling procedures described by Gickling and Havertape
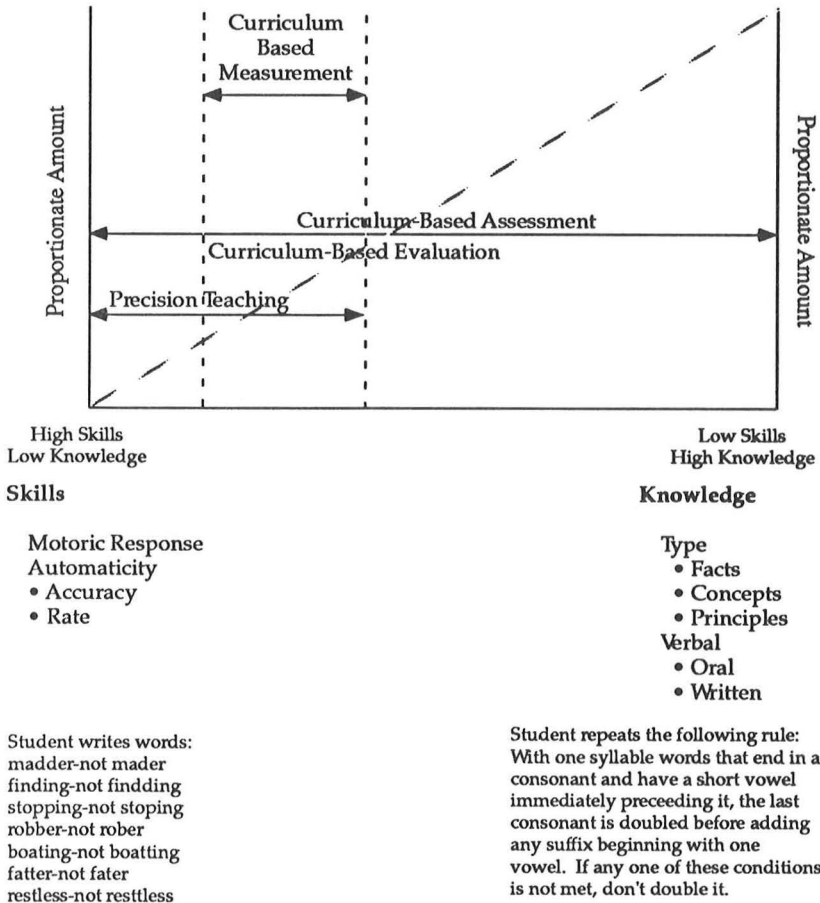
Figure 1. Skills versus knowledge focus of different models of curriculum-based procedures.

(1981) are actually curriculum construction techniques. The purpose of reading assessment is to find the ratio of known to unknown words and move the student from "unknown" to "known." In completing this goal, however, the balance of the ratio is critical, so procedures are described for developing reading passages with the appropriate blend of unknowns. The techniques for sampling items described by Idol, Nevin, and Paolucci-Whitcomb (1986) are based on criterion-referenced

test construction principles (defining a domain, sampling item types, and establishing mastery levels).

A major distinction between CBM and other forms of CBA revolves around consideration or definition of the curriculum. The curriculum is assumed to be an instructional variable like any other manipulatable variable. However, two issues must be resolved in developing a curriculum-based measure. First, the curriculum itself must be defined and second, alternate measures within that curriculum must be generated.

In many special education programs, a unique curriculum is used to instruct students in the basic skill areas. For example, Direct Instruction programs often employ Distar, Reading Mastery, Corrective Spelling, etc., in which not only teacher interactive strategies are highly specified, but the sequence of curricular materials is highly structured and organized. Using the long-range goal methodology suggested by Mirkin, Deno, Fuchs, Wesson, Tindal, Marston, and Kuehnle (1981), passages or word lists could be constructed from a wide band of units that reflect the year-long expectations. However, it is also possible to consider the curriculum used in the mainstream as the one from which measurement items should be sampled. For example, although a special education student may be receiving instruction in Corrective Spelling, alternate word lists could be developed from Kottmeyer, since the general education students are being taught and tested in that curriculum. This view of the curriculum is very broad and focuses on another important dimension of CBM that is reviewed later: a focus on the terminal response. Ideally, the behavior or skill that is being taught should not be curriculum bound, but should transfer across materials and settings.

A hallmark of curriculum-based measurement is the development of Individual Educational Plans (IEPs) using a long-range sampling plan, in which items are selected that will be taught within the academic year, but are not specific to the instructional levels on a daily basis. For example, Fuchs and Shinn (1989) and Mirkin, Fuchs, and Deno (1982) prescribe sampling reading passages, spelling words, or math computation problems for writing IEP goals that will appear within a student's lessons over the entire year. These items are then presented within a frequent measurement system that generates alternate forms that should be sensitive to student performance changes over time. The reading and math item-generation computer programs developed by Germann (1986a, 1986b) are simply tools that help teachers develop such alternate forms easily, by randomly sampling items from prespecified long-range goal domains.

To date, little research has been completed on this dimension of CBA, with most of it confined to research within curriculum-based measurement. Tindal, Marston, Deno, and Germann (1982) found differences between reading curricula in student oral reading fluency and speculated that it may be a function of the instructional emphasis of the curriculum (i.e., code versus meaning emphasis). Fuchs, Tindal, and Deno (1981) and Tindal and Deno (1981) sampled from domains of varying size and synchrony with instructional programs and found an intermediate level to be optimal for reflecting improvement over time with minimal variability; this level was neither as narrow as an instructional level nor as broad as a grade level. Fuchs, Fuchs, and Deno (1982, 1984) described the problems with varying passage readabilities that typically accompany a basal reading program and the implications for developing alternate forms within a curriculum-based measurement system. Finally, Shinn, Gleason, and Tindal (1989) analyzed two long-range goal sampling plans, one of which was near instructional levels and the other well beyond it; they found goals sampling from well beyond the instructional level to be sensitive to growth.

In summary, the curriculum is more broadly conceived in CBM than in other forms of CBA. The rationale is simply that sampling from units around (rather than within) the instructional level (which therefore includes preview and review of items) allows comparability across successive data values and is necessary for developing repeated measurements.

In the example below, a student being taught in a resource room using Reading Mastery was concurrently assessed in two curricula: (a) from instruction and (b) from the mainstream. In both systems, a long-range sampling plan was utilized, in which passages from within a 10-week period (one quarter) were selected randomly for each measurement. The only stipulation on this sampling plan was that no passage was allowed for measurement if it had been presented for instruction within 1 week. Because every passage had an equally likely chance of being selected, comparability of measures was possible. The question in this project was as follows: If a student is taught in a specialized curriculum, do the skills transfer to another curriculum? As reflected by the slope of improvement, general reading improvement is evident in both programs. However, the relative amount of improvement in the curriculum of instruction is greater than the amount of improvement in the generalized mainstream curriculum.
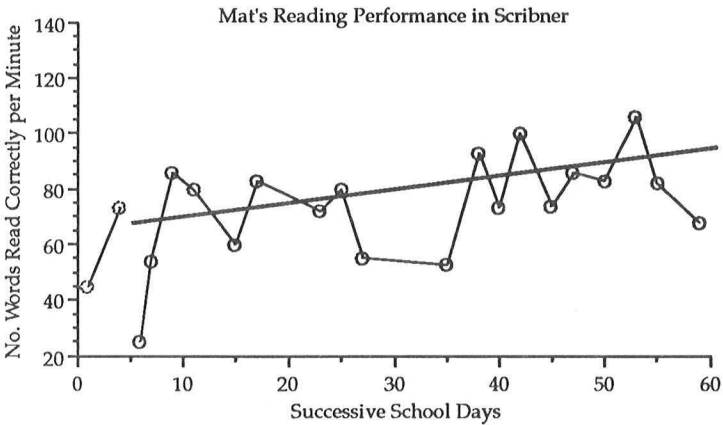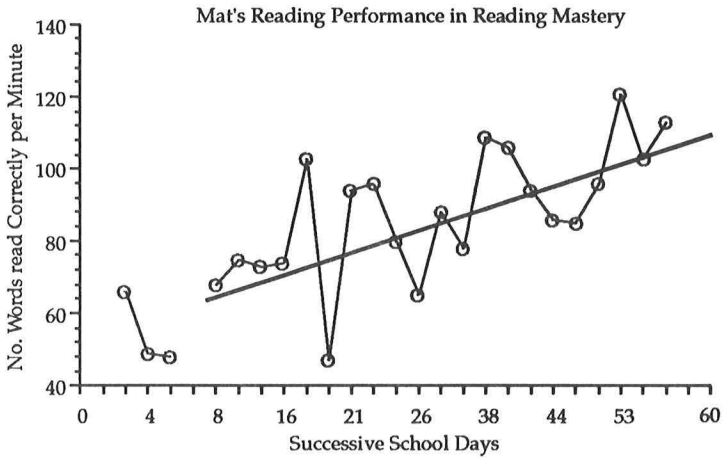
Figure 2. Student performance in the curriculum of instruction and the curriculum of the mainstream.

ADMINISTRATION AND SCORING PROCEDURES

To provide comparability in results, most assessment and measurement systems advocate using standardized administration and scoring procedures. Without constant directions, student

performance may be influenced inadvertently, either positively or negatively. Virtually all published measures of student performance, whether they focus on achievement, ability, perceptual processing, or latent traits, have explicit procedures, if not outright scripts, for test administrators to follow. Likewise, most advanced training programs in special education and school psychology devote a substantial portion of coursework to learning administration and scoring procedures with a variety of assessment devices.

This same dimension has important bearing in the area of curriculum-based assessment, evaluation, and measurement. The terms formal and informal can be used to characterize this dimension. Formal measurement systems employ standardized administration and scoring procedures, whereas informal measurement systems utilize nonstandardized techniques. These terms should not, however, be confused with published versus teacher-made, as is sometimes the case (Hargis, 1987). It is possible, and quite desirable, to have a measurement system that is teacher-made and formal (standardized); it is also possible (and quite undesirable) to have a published measure that is administered informally (in a nonstandardized manner), which is probably often the case in spring testing around the country. An example of a standardized administration procedure in reading is depicted in Figure 3.

Virtually all researchers of curriculum-based assessment and measurement have some description of administration and scoring procedures; some are simply more explicit than others. In Gickling and Havertape (1981), where analysis is predicted on the ratio of knowns to unknowns, the definition of an error is critical; yet, nowhere in the training module is information provided on how to administer a measure in reading or math (the only two areas covered) or how to score performance. For example, although the term "hesitant" is used to depict words that the student "near missed" in reading, it is uncertain whether such words represent those poorly decoded, self-corrected, or simply mispronounced using the wrong syllabication. In Figure 4, several published informal reading inventories are compared on how errors are defined, which can include any of the following: self-corrects, hesitants, assists, mispronunciation, omissions, insertions, repetitions, dialect, partial words, nonwords, substitutions, punctuations, and poor phrasing.

In contrast, the model proposed by Howell and Morehead (1987) is very explicit in the administration and scoring of curriculum-based evaluations. In fact, a major premise of their work is that the response itself is a very meaningful unit for diagnosis, and careful consideration must be given to definitions of errors and analysis of responses. The

work of Idol et al. (1986) also describes an explicit concern with administration and scoring issues. For example, in reading, they describe procedures for constructing 100 word sample passages and address the definitions of errors (omissions, substitutions, additions, and pauses are errors; repetitions, self-corrections, and deleted suffixes are not accounted as errors). Furthermore, all comprehension questions have prespecified answers used to score the students' response. Of course, in the multiple-choice responses, answers are keyed into the problems.

Materials Selection

Basal reading passages -- There are multiple selections for each grade representing a random sample from the grade level when used for norm-referenced purposes and Long Range Goals from the Individual Educational Plans, when used for individual-referenced purpose. Only passages that contain generally uninterrupted text (either expository or narrative) and devoid of unusual proper nouns, excessive dialogue, or poetry are included in the random sample. Each selection has a tester's copy (numbers on the tester's copy represent a cumulative count of words in the passage for each successive line. The length of the measurement is one minute.

Isolated word list -- There are different word lists for all grades, representing a random sample from the grade level when used for norm-referenced purposes and Long Range Goals from the Individual Educational Plans, when used for individual-referenced purpose. The measurement is conducted with two copies of each list -- one for the student to read from and a follow-along list for the tester to mark words read incorrectly. The follow-along list contains a cumulative count (by groups of 5 words). The length of each measurement is one minute.

Administrative Procedures

General directions. This test is individually administered and should be given in an area free from distraction. Put the student copies in front of and facing the student. Make sure they are in the same order as the tester's copies.

Take your copy, place an acetate sheet on top of it and put it in front of and facing yourself. Read the directions verbatim for the first administration.

When the student is finished, jot the score down, and quickly move to the next reading task; place the top sheet over and to the side and tell the student you would like to continue in the same manner. Repeat this procedure until all reading tasks are completed.

Specific directions. Say to the student: "When I say 'start', begin reading at the top of this page. If you wait on a word too long, I'll tell you to go on. If you come to a word you cannot read, just say 'pass' and go on to the next word. Do not attempt to read as fast as you can. This is not a 'speed reading' test. Read at a comfortable rate. At the end of one minute, I'll say 'stop'."

Scoring

Follow along on your copy, circling with a grease pencil incorrectly read words.

● Count as an error a misread word; i.e., house for horse, hug for huge, home for house, big for huge.
● Count as errors any words the student cannot read within about five seconds. After that period of time, tell the student to go on.
● Count an omission as an error. Count all words in skipped lines as errors.
● Count reversals as an error; i.e., saw for was.
● Do not count more than one error on the same word. For example, if the student mispronounces the same word more than once, count it only once.
● Do not count self-correction as an error.
● Do not count word additions or insertions.

At one minute, say "Stop." Place a slash after the last word read. Count the number of words read correctly and incorrectly.

For the basal reading passages, simply use the number to the right of the last full line read. Add to this the number of words read in the next (partially read) line. This represents the total number of words read. To obtain the number read correctly, subtract from this total amount the number of words read incorrectly.

Figure 3. Example of standardized administration procedures in reading.

Figure 4. Examples of different informal reading inventory error symbols.

Error Type	S-ARI	ARI	SRI	DRI	BRI	BRLI
Omission	He had lost	Had had (lost)	He had lost	He had lost	He had (lost)	He ~~had~~ lost
Insertion	*had* I ∧won	*easily* I∧won	*big.* I won∧	*shouldve* I ∧won	*quickly* I∧won	*couldve* I∧won
Substitution	*pantry* A big ~~party~~	*The* A big party	*large* A ~~big~~ party	*busted* A ~~big~~ party	*The* A big party	*pantry* A big party [1]
Mispronunciation	*been* She'd been out					She'd been out
Nonpronunciation	Really bought [1]					
Aid/Assist		They rested	They (rested)	They res?ed	*P* They rested	*P* They rested
Repetition	✓The latest	The latest [1]			The latest [1]	*R* The latest
Reversal		(Tired, old feet			(Tired, old feet	
Hesitation		Starved /for food				Starved for food
Non-Errors Repetition			*R* The latest	The latest		
Self-Correct	*Sc* He wasn't sure	*sc sur* He wasn't sure	*Sc* He wasn't sure	*(was.* He wasn't sure		*S sur* He wasn't sure
Poor Phrasing	He ran home					He/ran/home
Hesitation			*H* He ran home	He ran home		He ran home
Partial Word					*des-* Destitute	
Non-words					*gone* He hadn't gone	
Punctuation		*P* I won't. But she...				I won't␘ But she...
Dialect					He was talkin' loud	He was talkin' loud
Mispronunciation (Intonation)		*M* Very plentiful			Very plentiful	

[1] five seconds | [1] two or more | | | [1] Corrections vs. repetitions | [1] Disrupt meaning Same meaning

TYPE OF RESPONSE

An important issue in any measurement-testing system is the type of response that is generated by the person taking the test. Two types of responses are possible (Hopkins & Antes, 1978): production or selection.

A major component of all curriculum-based measurement research is attention to administration directions and scoring procedures. In Figure 3, an example of the procedures generally followed in reading is provided. Note that the standardization process includes procedures for sampling and formatting materials, administering the measures, and scoring the responses. Since production responses are generated, scoring procedures that utilize objective rather than subjective criteria are critical.

In a production response, the examinee actually constructs or produces the answer, which is then scored for correctness or quality. Generally, three types of responses can be made (Tindal & Marston, 1990): (a) one word, an example of which is the cloze format (McKenna & Robinson, 1980); (b) short answer, which is often employed in informal reading inventories; and (c) extended answer, used in the traditional essay examination in high schools and colleges.

In a selection response, the examinee is provided the test stimulus and a range of options or answers, only one of which is correct and should be selected. The basic form is multiple-choice, which can be formatted (a) with the traditional four or five options, (b) as a true-false proposition, or (c) as a classification-matching problem. Virtually all published achievement tests, both norm and criterion referenced, employ selection responses. By having the examinee fill in a bubble on an answer sheet, it is possible to group administer and computer score the test, both which create a cost-efficacious measurement program. However, some newer achievement measures are being constructed with production responses, most of which are marketed through PRO-ED, Inc. (i.e., Test of Written Language-2, Test of Written Spelling, etc.).

The selection and production dimensions also provide an interesting focus for analyzing curriculum-based assessment and measurement systems. The model proposed by Idol et al. (1986) broadly encompasses both types of responses. In reading, oral and silent responses are considered, with comprehension assessed using a question-answering format, both oral and written. The model of curriculum-based evaluation proposed by Howell and Morehead (1987) also includes both response types. Oral reading and decoding primarily employs a production response, whereas comprehension is assessed using a variety of

procedures: cloze and retelling (production), and maze and multiple-choice answers to questions (selection). Gickling and Havertape (1981) focus on both production and selection responses through their examples in the training module. Students orally read, compute answers to math problems, spell words (all of which represent production responses), and select the correct word to complete sentences.

In contrast to curriculum-based assessment, all examples of academic assessment reported in the precision teaching journal are production responses. In general, the research on curriculum-based measurement is limited to production responses, with the exception of the maze task in reading. The behavior of focus in reading is oral reading from passages and word lists (Deno, Mirkin, & Chiang, 1982), with oral and written responses (number of words produced) that "retell" the content from passages (Fuchs, Fuchs, & Maxwell, 1988; Krauss, 1988; Tindal & Parker, in press). In mathematics, responses have been confined to completion of computation problems (Tindal, Germann, & Deno, 1983). Spelling measurement has been limited to two types of production responses: words spelled correctly and correct letter sequences (Deno, Mirkin, Lowry, & Kuehnle, 1980). In written expression, a number of different responses have been investigated, all of which are based on an analysis of the student's composition and therefore are production responses, including the number of words written, words spelled correctly (Deno, Mirkin, & Marston, 1980), and words in correct sequence (Videen, Deno, & Marston, 1982). In recent research completed on reading group placement, Parker, Hasbrouck, and Tindal (1989) used a maze test in reading. Presently, no other responses have been investigated in content areas outside these basic skill areas. Consistent with this orientation on production responses, Shapiro (1989) includes many of these responses just noted in his book on academic skill assessment.

TECHNICAL ADEQUACY

Any measurement system must be reliable and valid to be used in making decisions about students. This concern with reliability and validity is not limited in its application to formal, published achievement measures; rather, all measures of achievement eventually must have established technical adequacy, whether developed by curricular publishers or individual teachers. Likewise, curriculum-based assessment, evaluation, or measurement must be analyzed first and foremost by established test standards developed and promulgated by

the American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (1985).

Using the classical definitions of these terms from Nunnally (1967), reliability, or consistency, is considered necessary but not sufficient for validity or truthfulness. Reliability is further organized into four different types, according to the source of potential error: test-retest, alternate forms, split-half, and interjudge. Validity is further refined into four different types: content, concurrent, predictive, and construct. In applying these concepts to the research on CBA/CBM, it is clear that suggestions for measurement often overwhelm and precede any supporting data. Simply stated, very few technical adequacy data have been generated by the proponents of curriculum-based assessment. In contrast, scores of studies have been completed on various aspects of the technical adequacy of curriculum-based measurement. Rather than focus on the lack of information for the various versions of CBA, the remainder of this section will simply highlight the major findings on CBM that have appeared in the published literature.

1. Fuchs, Fuchs, and Deno (1982) analyzed the reliability and validity of CBM oral reading measures and found them to be both reliable and criterion valid with respect to the Woodcock Reading Mastery Tests and teacher judgment.

2. Deno, Marston, Shinn, and Tindal (1983) reported on differences in oral reading performance among students of different grades and classifications.

3. Deno, Marston, Mirkin, Lowry, Sindelar, and Jenkins (1982) studied the developmental trends of reading, writing, and spelling performance over the grades at different levels of proficiency and established interestingly regular growth curves.

4. Deno, Mirkin, and Marston (1980) investigated the criterion validity of the number of words and words spelled correctly in response to a story starter and found moderately high correlations with the Test of Written Language.

5. Deno, Mirkin, and Chiang (1982) found very high correlations between the number of words a student could read orally in 1 minute and their performance on different subtests from published reading achievement measures.

6. Deno, Mirkin, Lowry, and Kuehnle (1980) found that students' proficiency in spelling words correctly and concatenating letters in correct sequence was related highly to their performance on spelling subtests of published measures of achievement.

7. Fuchs, Deno, and Marston (1983) analyzed the reliability of curriculum-based measures as a function of the duration of behavior sampled and found 1 minute to be adequate, with longer times producing more consistent results.
8. Fuchs, Fuchs, and Maxwell (1988) analyzed the criterion validity of oral reading and retelling with the Stanford Achievement Test and found moderately high correlations.
9. Marston and Deno (1981) researched the reliability of the written expression measures (using the number of words written and spelled correctly), and Tindal, Marston, and Deno (1983) expanded this study of reliability to measures of reading, spelling, and math.
10. Shinn, Tindal, and Stein (1988) summarized the research on the use of curriculum-based measures in differentiating students labeled low achieving and those classified as learning disabled.
11. Tindal, Fuchs, Fuchs, Shinn, Deno, and Germann (1985) compared several curriculum-embedded mastery tests with CBM and found moderate relationships, which were limited because of the low reliability of the mastery tests.

In summary, many different studies have been completed on the technical adequacy of CBM, with most of the data very supportive. This research has been conducted in several parts of the country, with students from many different grade levels and ability groups, using a variety of methodologies and many different criterion measures (i.e., a variety of achievement tests, both criterion and norm referenced; teacher judgment; classification differences; age differences; and growth over time). Although more research needs to be completed on the technical adequacy of CBM, the data that have been generated should outweigh the criticisms by skeptics proposing other systems for which no data have been generated.

FREQUENCY OF MEASUREMENT

An important dimension for evaluating measurement procedures is their frequency of administration. Most norm- and criterion-referenced tests are designed for single administrations; most behavioral measures are individually referenced, with repeated measurement allowing comparisons of current levels and rates of performance changes to previous levels and rates.

The difference in administration frequency is not a slight matter, but represents a fundamental difference in the basic datum for reflecting student performance. With a norm-referenced measure, all scores are

related to the position of the individual within the group. For example, standard scores, percentile ranks, and normal curve equivalent scores are all transformations of an individual's raw score relative to others' scores on the same measure. With most criterion-referenced measures, the score represents an absolute level below which is failure, noncompetence, or nonmastery and above which represents success, minimal competence, or mastery. Although this cutoff may be established using any one of several methods (Berk, 1986) and may include an error term for analyzing classification accuracy the cutoff eventually reduces the outcome to one of two possible states.

With a repeated measurement approach, which is an underpinning of a behavioral perspective (Tawney & Gast, 1984), the datum for summarizing performance is change over time or slope of improvement. For deficit behaviors, in which growth is expected to increase (i.e., reading fluency), a positive and steep slope is desirable; for excess behaviors (i.e., hitting), the goal of interventions is to generate a negative and steep slope. Another dimension that is available with frequent measurement is the individual variation across successive measures. Finally, overlap, or the percentage of data values within the same range (Scruggs, Mastropieri, & Castro, 1987; White, 1987), provides a metric for quantifying changes in performance over time. Together, these three indices can be used as a datum for describing performance. As Parsonson and Baer (1978) note, they can be used within and across instructional phases, generating a very rich and complex data base for evaluating student performance.

Frequency of measurement simply has not been addressed explicitly in the professional literature on curriculum-based assessment. Some researchers have described systems which lend themselves well to a specific datum; however, no explicit research has been completed in this area. The datum used by Gickling and Havertape (1981), reflecting the ratio of known to unknown items on well-specified domains, appears to be oriented around a criterion reference; the literature on active learning time, which provides the rationale for their outcome metric, suggests high levels (at least 90%) of success for learning to be optimal. The model proposed by Idol et al. (1986) also appears to have a criterion reference, since mastery states on explicitly defined tasks are proposed. Howell and Morehead (1987), in using a "criterion for acceptable performance" on specific level tasks (well-defined domains) provide yet another example of a criterion reference. In all these examples, repeated measurement is not generally emphasized. Rather, post-only or pre-post measurement is employed.

Repeated measurement appears as a central tenet with only two models: precision teaching and curriculum-based measurement (Tindal & Germann, 1985). However, only a few studies have been completed. Tindal (1983) investigated the reactivity of outcome judgments to changes in slope, level, and variability and found teachers differentially consistent. At times, slope appeared dominant in the judgment process and at other times, variability in performance was the major datum for assaying outcomes. Skiba, Marston, Wesson, Sevcik, and Deno (1983) analyzed the characteristics of time series data upon which CBM is predicated.

Because most of the research on data utilization is premised upon a frequent measurement model (Tindal, 1988), it is not possible to isolate its effects apart from the manner in which data are used to formatively evaluate instructional programs. However, in a meta-analysis on the effects of systematic formative evaluation, in which data utilization was randomly confounded, Fuchs and Fuchs (1986) reported very impressive outcomes. When teachers measure students frequently and graph performance, an effect size (Hedges and Olkin, 1985) of .25 was present. In a similar vein, Marston, Fuchs, and Deno (1985) compared published achievement measures (norm referenced) with curriculum-based measures (individual referenced) and found the latter to be more sensitive in reflecting changes over time. It is uncertain whether this differential sensitivity is a result of the curriculum-specific sampling plan or different metric using frequent measurement for summarizing outcomes. Finally, in an interesting focus on evaluation methodology, Marston (1988a) used a time-series analysis to assay the effectiveness of special education. Arguing that the appropriate control comparision for special education is not peers from a normative standardization sample, but rather previous performance prior to special education, he used an AB (regular-special education comparison) to determine whether the slope of performance change was greater in special education. His results confirmed this prediction. In summary, an essential feature of CBM has been the use of frequent, time-series or repeated measures, with some empirical justification for its consideration.

DISPLAY AND ANALYSIS OF DATA

Eventually, all measurement and assessment data must be analyzed, displayed, and interpreted. Current technological innovations in computers create many impressive options for completing the operations. Few schools operate without computers in the classroom

and these computers are being used at ever-increasing rates to handle the mundane tasks of "number crunching" as applied to assessment. However, special education applications of computer technology within the assessment process has been confined generally to report writing and IEP management (Enell & Barrick, 1983; Jenkins, 1987; Ryan & Rucker, 1986).

The issue of data display has not been addressed by most researchers investigating curriculum-based assessment; however, it is a very important component of curriculum-based measurement and precision teaching. Generally, graphic display of data has been considered instrumental in data utilization, with primary emphasis on line graphs (Tindal, 1987). Research conducted on CBM has been confined to equal interval graphs, while the research completed on precision teaching has utilized logarithmic graphs, typically using six cycles and known as the Standard Behavior Chart. The biggest problem, however, has the polemics which appear from both sides, often precluding a rational or empirical analysis. One of the few studies to be completed on the type of graph was reported by Marston (1988b) and Marston and Deno (1982); they found equal interval graphs to have higher accuracy in predicting student performance over a 2-week period.

In the reseach on graphic displays of student performance and data utilization, a number of issues have been addressed, including frequency of measurement (Mirkin, Deno, Tindal, & Kuehnle, 1980), types of decision rules that accompany graphic displays of data (Mirkin & Deno, 1979), formative evaluation of instructional programs (Tindal, 1988), and graphic factors (e.g., slope and variability) influencing judgments and interpretations (Tindal & Deno, 1983).

This research has not been confined to simple progress charts of individual students, but has also focused on analyses of normative distributions. Given the multi-decision focus on CBM, in which screening and eligibility are an important component, normal distributions are critical for valid decision making. For example, if the distribution of a group of first graders, obtained in the fall of the year, is leptokurtic and positively skewed (a very likely event), it is difficult to make valid decisions about low-achieving students. Most students in first grade have few basic skills. Therefore, in the analysis of normative displays (Shinn, 1988; Tindal, Germann, & Deno, 1983), the shape of the distribution and its "normality" have been emphasized. In the figures below, two radically different distributions have been obtained on CBM-like measures, with the first one non-normal (a writing task completed by low-achieving and remedial first graders in the fall) and the second one very normal (a reading task completed by general education fourth-grade students in the fall). ✓
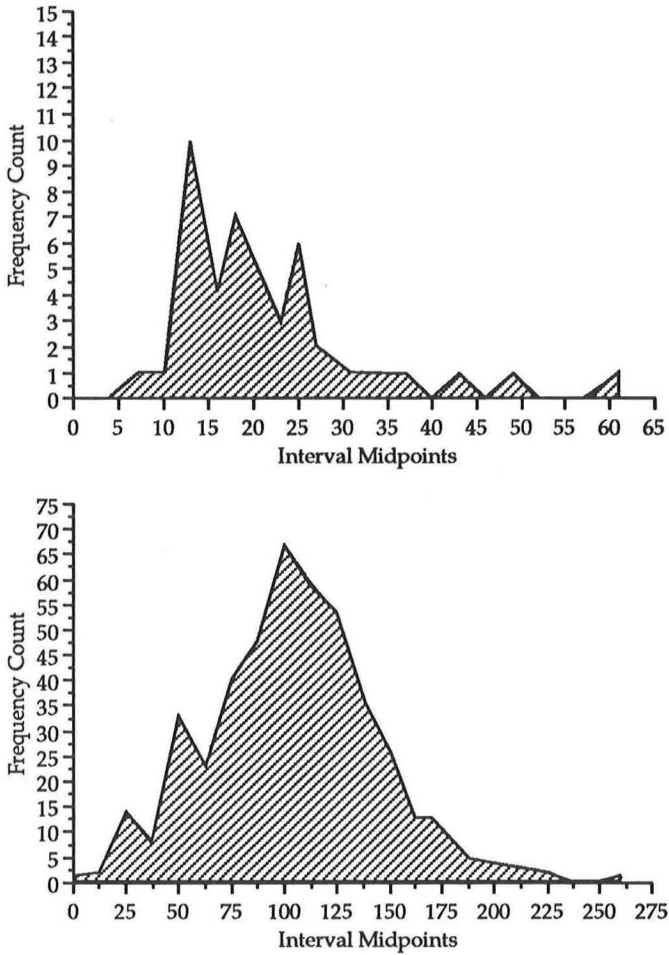
Figure 5. Graphic displays of data: Norm referenced distributions on oral reading flueucy.

REFERENCE GUIDES FOR DATA INTERPRETATION

When students are tested and measured, two interpretive judgments can be made: one that focuses on process (how students perform) and the other that addresses product or outcome (how well students perform). Generally, this outcome is a number of some type (i.e., is based on an ordinal or interval scale). However, the number itself is quite uninterpretable without a reference with which to anchor it. Three different types of references can be used to provide meaning to student test outcomes.

Norm-referenced testing. In this reference type, students are compared to each other on a commonly administered and scored measure. Often, the term is inappropriately considered synonymous with published tests and/or contrasted with teacher-made tests. However, it is possible to devise a test that is norm-referenced and not published (i.e., many curriculum-based measures are norm referenced and not marketed); it is also possible to have teacher-made tests that are norm referenced. The other point of confusion frequently made with the two terms is between *norm referenced* and *standardized*. Although norm-referenced tests must have a sample of students upon which the norms are based, often referred to as the standardization sample, the test may be administered and scored in either a standardized or nonstandardized fashion.

Because norm-referenced tests employ comparisons of students to each other in the interpretation of performance, the composition and comparability of the student group is critical. Although this issue may seem obvious, many tests are published that have very limited norms (Conoley & Kramer, 1989; Mitchell, 1985), and as Ysseldyke and Thurlow (1984) note, these tests are nevertheless commonly used to make many important educational decisions. Reviews of several commonly used norm-referenced measures appear in Salvia and Ysseldyke (1988) and Witt, Elliott, Gresham, and Kramer (1988).

In a norm-referenced interpretation, a student's relative position in a distribution is the most important interpretive index. The average performance and the amount of variability in the group are used to index this position. Interpretations using norm-referenced guides are generally based on frequencies and probabilities. For example, a student with a score of 55 on a test with an average score of 50 and a standard deviation of 10 is considered average, since the score is at a position on the distribution with many other scores. In contrast, a student with a score of 15 on this same test would be very deviant, since this score is at a position in which very few scores lie.

A host of different score transformations can be made with these three pieces of information. For example, performance may be reported in standard score units of several different types, using an interval scale (e.g., z-scores and T-scores), a pseudo-interval system (age-grade equivalent scores, which are not recommended), or a ranking system (i.e., percentile ranks and stanines). Although these scores differ in the information conveyed, they all reflect the student's relative position in a distribution.

Given these overall qualifiers, few curriculum-based assessment systems have been developed or reported in the professional literature. In contrast, a number of studies have appeared in which curriculum-based measurement is used in a norm-referenced manner. For example, Shinn (1988) describes how norms can be generated and utilized in decision making. Tindal, Germann, and Deno (1983) reported on several technical characteristics of the norms that were generated in the Pine County, Minnesota, Special Education Cooperative. Tindal, Shinn, and Germann (1987) used a norm-referenced approach in evaluating special education effectiveness and found differential sensitivity in the different score summary systems. Finally, in the many studies on screening and eligibility reported in the section of decision making, a norm-referenced approach has been used (Shinn, Tindal, & Stein, 1988).

Criterion-referenced testing. The general definition of this interpretive reference is that (a) a specific domain of items is identified and (b) a sampling plan for selecting these items is operationalized. In most systems, a criterion for mastery is also defined (Popham, 1984). Although not requisite to a criterion-referenced approach, mastery status has functionally been intertwined with the definition of criterion referencing (i.e., a domain may be established without mastery, though mastery implies that a specific domain has been identified). Many books have been written that specifically detail procedures for developing criterion-referenced tests (i.e., Roid & Haladyna, 1982; Carey, 1988; Ebel & Frisbie, 1986) with the general focus on defining an appropriate universe of instruction from which to sample student learning. The technology of test construction is generally quite straightforward and noncontroversial, with a variety of procedures available (e.g., using selection or production responses, defining domains that are sequentially or hierarchically ordered, using different sampling plans). The real controversy in criterion-referenced testing comes from the establishment of mastery (Glass, 1980; Popham, 1978). In part, the problems arise from technical issues (Hambleton & Swaminathan, 1978). However, problems

in defining mastery are also a function of the judgmental nature of the process (Berk, 1986; Livingston & Zieky, 1982).

Most curriculum-based assessment systems are criterion referenced, with well-defined domains and established levels of mastery. For example, the procedures outlined by Idol et al. (1986) very specifically detail strategies for organizing a domain of instruction and developing a mastery level. Howell and Morehead (1987) also describe specific level assessment, a form of domain definition that is very hierarchically ordered, and "criteria of acceptable performance," a level of mastery status. The model of CBA proposed by Blankenship (1985) is also very consistent with this approach. Her description of formatting a CBA includes listing skills that are taught in the curriculum, organizing them into broader goals and objectives, which are in turn used to structure test items and generate student responses:

> Give the CBA immediately prior to beginning instruction on a topic. . . . Readminister the CBA after instruction on the topic. Study the results to determine: Which students have mastered the skills and are ready to begin instruction on a new topic. . . Periodically re-administer the CBA throughout the year to assess for long-term retention. (p. 234)

All models of CBA appear very closely aligned with criterion-referenced testing in their definitions of specific domain, strategies for selecting items from those domains, and particularly in establishing levels of mastery that are used to control progress through a curriculum.

In contrast, curriculum-based measurement includes mastery in the development of IEPs, but emphasizes individual referenced evaluations, as discussed in the next section. The work that has been done on the use of mastery states, though, provides some interesting findings that highlight its importance (Fuchs & Fuchs, 1984). Similar to the models of CBA described above, the criterion-referenced perspective focuses on three issues: the conditions under which the student is expected to perform, the behavior that is to be displayed, and the level of proficiency that is needed. Mirkin, Deno, Fuchs, Wesson, Tindal, Marston, and Kuehnle (1981) describe the procedures for completing IEPs in the basic skill areas, employing these three components. However, rather than arranging skill areas within well-delineated domains that are sequenced hierarchically, the domains that are defined within a CBM approach are diverse and include many subskill areas (Fuchs, Tindal, & Deno, 1981; Tindal & Deno, 1981). The long-range goal that is specified within an IEP, therefore, literally reflects the domain that the student is expected to master by the end of the monitoring period, usually an academic year (Fuchs & Shinn, 1989). Although the materials from within this domain are then randomly

selected, the initial definition of the domain is far from randomly determined. Tindal (1984) describes several procedures for establishing an appropriate domain, using (a) student performance across different levels of material, (b) standards appearing in the professional literature, (c) normative performance on standard tasks, and (d) expert judgment of the teacher.

Given this contrast in defining domains between CBA and CBM, definitions of mastery assume a very different meaning, with more emphasis on progress toward mastery rather than actual attainment of mastery. However, as Fuchs, Fuchs, and Deno (1985) have demonstrated, the expectations (absolute levels of mastery on broadly conceived domains) are extremely influential on eventual attainment of proficiency (see also Fuchs, this volume).

Individual-referenced testing. In the previous approaches, the standards used to interpret student performance are externally derived, either through peers' performance or some judgmental process. In an individually-referenced approach, the *progress* of the student is most important; therefore, the standards become rate of change over time, which is internally derived. Using a single subject methodology (Tawney & Gast, 1984), slope of improvement replaces levels of proficiency as the basic datum for evaluating programs.

To develop this frame of reference, however, requires that an appropriate domain definition and sampling plan be available for generating comparable alternate forms of measurement over time. Every data point needs to be comparable to all other data points; this provides the basic rationale for random sampling on long-range goal material in the IEPs. If every item has an equal probability of appearing on a single measure, and the items both preview and review material, comparability is achieved in the measures used for monitoring progress. However, because any one measure actually may be different from another one, the level of performance on the measures is replaced with the slope of improvement across the measures. In many of the graphs that have been generated in both research and practice using this technique, variability indeed is apparent, reflecting a domain or sampling effect (see Fuchs, Fuchs, & Deno, 1982, 1984 for a review of the issues in sampling passages with varying readabilities).

The models of CBA generally cannot be used in a time-series format, other than to display mastery of successive units (see Deno & Mirkin, 1977 for a description of mastery monitoring). In contrast, the research and practice appearing with CBM is replete with data using an individual-referenced approach. Generally, one of two approaches has been used to organize such evaluations: treatment or goal oriented

(Deno & Fuchs, 1987; Fuchs, 1986). In the former, evaluation focuses on the treatments, using an ABCD design (Tawney & Gast, 1984) in which successive treatments are compared to each other in determining which one is the most effective. This design was used in providing the data base reported by Casey, Deno, Marston, and Skiba (1988) in an experimental teaching project and by Deno, Chiang, Tindal, and Blackburn (1979) in a program evaluation. In contrast, the latter technique uses IEP goal attainment to help structure the evaluation process. This procedure appears less frequently in the published literature, but probably is more widespread in CBM implementation sites (i.e., Pine County, Minneapolis). Tindal (1988) summarizes the literature on individual-referenced evaluations, including these two techniques (treatment- and goal-oriented foci) and the use of long- and short-terms goals to structure the outcomes (Fuchs & Fuchs, 1986; Tindal, Fuchs, Christenson, Mirkin, & Deno, 1981). These two procedures are illustrated in Figure 6 below:
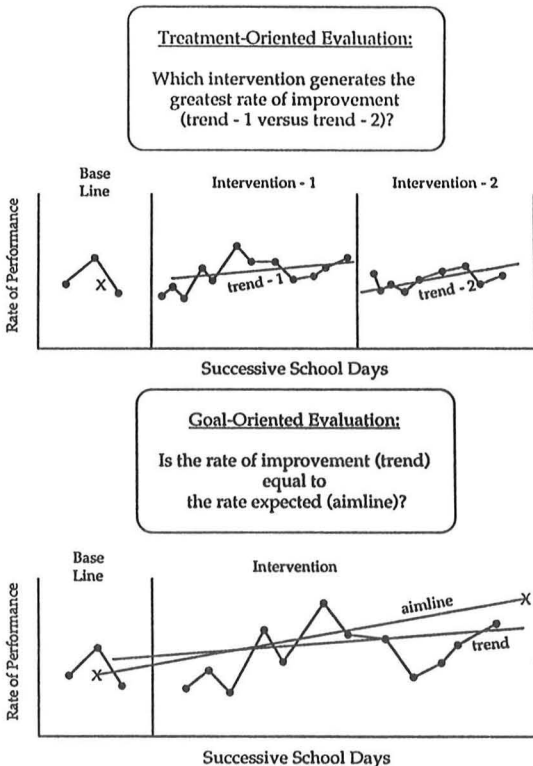
Figure 6. Two types of individual-referenced decisions: Treatment and goal oriented evaluation strategies.

USE IN DECISION MAKING

Although different types of educational decisions have been identified in the professional literature (Deno & Mirkin, 1977; Salvia & Ysseldyke, 1988), these decisions generally revolve around three major functions: (a) allocation of resources (screening and eligibility), (b) instruction (planning and evaluation of methods and materials), and (c) evaluation of programs (Posavac & Carey, 1989). Generally, norm-referenced data are used to make screening/eligibility decisions and to evaluate overall program outcomes, whereas criterion- or individual-referenced data are used to plan and evaluate instruction (Tindal & Marston, 1990). A depiction of this differential use of data for specific decision making is presented in Figure 7.

Decision	Screening Eligibility	Program Evaluation	Instructional Planning Formative Evaluation	
Reference Type	Norm-referenced		Criterion-referenced	Individual-referenced
Curriculum Sampling	Multi-curricula Multi-unit		Mono-curricula Mono-unit	Mono-curricula Multi-unit
Data Scale	Continuous		Discrete	Continuous
Threats to Interpretation	Underlap with curriculum Sensitivity to growth Limited behavior sampling Normative group composition		Noncomparability across units Generalization & maintenance Limited behavior sampling Determination of mastery	Sampling Domain Sensitivity to grow Outcome metrics

Figure 7. Characteristics of the data base and reference type for different educational decisions.

Program decisions (screening/eligibility and program evaluation) tend to use norm-referenced data because of the need to generate comparable measures for many individuals over an extended time period; such data can be considered broad band with low fidelity. In contrast, instructional decisions need to be specific to individual students over a more limited time period; these data are narrow band with high fidelity.

The band width is determined in great part by the curriculum-sampling plan. Norm-referenced data typically sample from across several curricula (and across several units within a curriculum). This

aspect of their construction has led many authors to assert that they have little content validity (Freeman, Kuhs, Porter, Floden, Schmidt, & Schwille, 1983; Good & Salvia, 1988; Jenkins & Pany, 1978; Leinhardt & Seewald, 1981; Shapiro & Derr, 1987). This may be less a problem of their construction than their use, however (Messick, 1981). Most norm-referenced measures have at most two alternate forms, generating pre/post measures, which rely on a continuous scale of change. Together, this broad sampling and minimal administration create some limitations in the interpretations that can be made from the data. A minimum range of item types are present that may not include the full range utilized within instruction; this problem may, in turn, limit the sensitivity of the measure to reflecting growth. Since measurement generally occurs only once or twice per year and within a concentrated administrative setting (i.e., one 45-minute period), the behavior that is sampled may be further limited. Because norm-referenced measures attain their meaning through the use of score transformations using a normative group, all measures of change are limited by the comparability of the norm group. Finally, the outcome metric may be more or less sensitive in reflecting change in student performance (Tindal, Shinn, & Germann, 1987).

Instructional decisions (planning and formative evaluation), given their greater specificity to individual students, must be confined to a specific curriculum. As presented in the section on curriculum sampling, differences exist, however, in the definition of curricula and the inclusion of material within or across instructional episodes; hence, the two options of either criterion or individual referencing. In the former, sampling is limited to within units, whereas the latter implies sampling across units. This feature, in turn, results in two different types of scales for summarizing behavior: a discrete one with criterion-referenced measures (Deno & Mirkin, 1977) or a continuous one with individual-referenced measures (Skiba & Deno, 1983).

Both approaches cited above contain several interpretive threats. The biggest problem with criterion-referenced measures involves the potential for differential difficulty and discrimination from one test to the next without very careful planning and development of test specifications (Carey, 1988). Since these measures are isomorphic with instruction, assessment results may be inaccurate after a period of noninstruction; generalization and maintenance may, therefore, be suspect. Generally, item types are minimally represented, presenting the same problem that appears with norm-referenced measures. Finally, mastery is essentially a judgmental process that is always in need of justification (Livingston & Zieky, 1982).

Individual-referenced measures also have limitations, mostly revolving around the definition of sampling domains (their breadth and the item selection techniques); in turn, sensitivity to growth may be differentially influenced by the sampling plan (Tindal & Deno, 1981). Finally, with a wide range of outcome metrics possible (i.e., slope, variability, step, overlap), assessment of change may be a function of the metric employed (Skiba, Deno, Marston, & Casey, 1989; Tindal, Deno, & Ysseldyke, 1983).

Virtually all models of CBA use a criterion-referenced approach to measurement and, as a consequence, focus on instructional planning and formative evaluation. For example, Idol et al. (1986) note that "curriculum-based assessments are teacher-constructed tests designed to measure directly students' skill achievement at specified grades. The assessments are criterion-referenced, and their content reflects the curricula used in general education classrooms" (p. v). Similarly, Tucker (1985) writes that "curriculum-based assessment is the ultimate in 'teaching the test,' because the materials used to assess progress are always drawn directly from the course of study" (p. 200). Other models proposed by Blankenship (1985) and Rosenfield (1987) also focus on instructional decisions; such measures are less useful at the program level.

Curriculum-based evaluation (CBE) (Howell & Morehead, 1987) and curriculum-based measurement (Deno, 1985; 1989), in contrast, span a range of educational decisions, including the instructional focus noted above and both program level decisions: screening and eligibility (allocation of resources) and program evaluation. CBE specifically describes a model of assessment that moves from survey level to specific level; the former term is clearly oriented around a broad sampling plan of items that may be very appropriate for screening students and evaluating outcomes across students and over time. The research on CBM likewise includes many different studies at each decision focus. Shinn, Tindal, and Stein (1988) summarize the research that has been conducted with the use of CBM to screen students and identify them as eligible for specialized programs. Tindal (1988) summarizes the research on instructional decision making, which primarily focuses on formative evaluation, rather than the instructional planning that is covered in the specific level assessments of Howell and Morehead (1987). Finally, program evaluation research is described by Tindal (1989), in which all three references (norm-, criterion-, and individual-referenced strategies) have been used to evaluate large-scale programs.

SUMMARY: A FINAL COMPARISON OF MODELS

Different models of CBA have been compared on a number of assessment and measurement features. The differences are striking on some of these features and quite minimal on others. For example, virtually all models begin with the premise that measurement items need to sample from the curriculum; the differences arise in how that curriculum is defined. The use of production versus selection responses may actually represent a minor variation that is not important among the models of CBA, CBE, and CBM, since they all include items of each type. However, the production/selection distinction is important in differentiating these approaches from most published achievement measures. Likewise, the focus on basic skills appears in all models; the extension of measurement into the content areas is simply more developed in a few curriculum-based procedures. It is possible that graphic displays could be incorporated into all models of CBA, CBM, and CBE; however, it appears to be a major emphasis of CBM and Precision Teaching. Finally, the use of standardized administration and scoring procedures could also become a major component of any one model; it is overtly emphasized (prescribed), however, in one application of CBA (Idol et al., 1986), CBE (Howell & Morehead, 1987), and CBM (Shinn, 1989).

The most fundamental differences appear to be on three features. First, let us consider the research on technical adequacy. Although the models and procedures other than CBM contain many very sensible ideas that are instructionally focused, little data are available to support them. The only exception may be the CBE procedures offered by Howell and Morehead (1987), which are built on a considerable diagnostics research base. However, the work of Gickling and Havertape (1981) and Gickling and Thompson (1985), which is further advanced by Rosenfield (1987) and Tucker (1985), has very little data supporting it. The models presented by Idol et al. (1986), although following best practices in test construction, simply have not been deployed in any active research programs. Bursuck and Lessen (1985, 1987) and Shapiro (1989) follow many of the procedures used in CBM.

Second, both the datum for summarizing student performance and its reference appear considerably different across the various models. CBA is oriented toward accuracy of performance and is criterion referenced. In contrast, CBM is oriented toward rate of performance and is referenced to norms, criteria, and individuals. Finally, CBE focuses on both accuracy and rate and is referenced from both norms and criterion domains and standards. Underlying this distinction is an

emphasis on domain definition, which can be either broadly gauged and useful for many different individuals and over time, or finely focused and applicable for specific individuals within a relatively short time period. In Figure 8 below, this feature is defined as item sampling, which can vary on a continuum from locked (tests and instruction are isomorphic) to linked (test items are sampled from instruction) to unrelated (with generic problems that may be similar in format but not content).

The above distinction is highly related to the third and final feature, the decision for which the data are employed. With a criterion-referenced focus, the major decisions center on instruction; in contrast, a norm-referenced focus clearly is appropriate for allocating resources and evaluating programs. Individual-referenced decisions, though designed specifically for instructional planning and evaluation, can also be used to allocate resources (Marston, 1988) and evaluate programs (Marston, 1987). These major decisions are organized on a continuum displayed in Figure 8 below. On one end are screening and eligibility decisions (allocation of resources), which can also include program evaluation; the next decision involves instructional planning and diagnostics; finally, instructional evaluation is the last major decision.
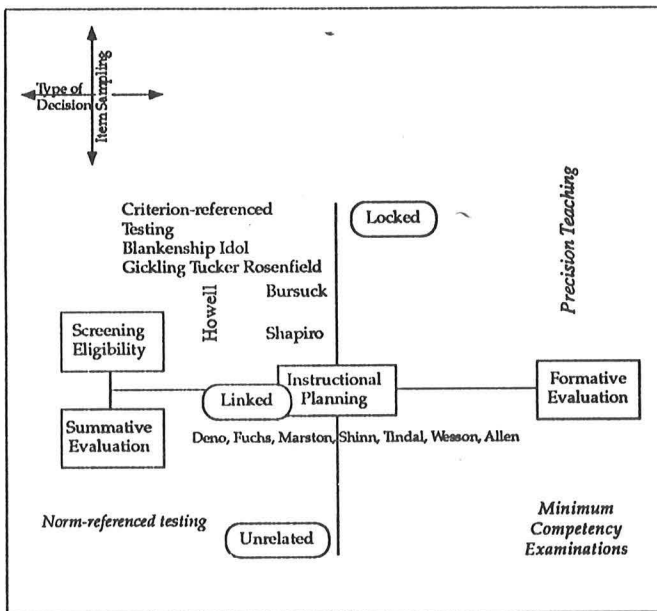
Figure 8. Comparison of different models of curriculum-based procedures on item sampling and type of educational decision.

In Figure 8, the various models of CBA, CBE, and CBM are compared. The major authors who write about them are located at an intersection relating the item sampling and type of decisions. Three types of testing that contain so many individuals are depicted without authors: norm-referenced, minimum-competency, and criterion-referenced testing. Likewise, given the number of individuals engaged in and the general dearth of published literature regarding precision teaching, the generic form has been used without specific reference to any individual authors.

In summary, the nine components discussed herein not only define curriculum-based procedures, but also provide educators with criteria for evaluating them and adopting them in their schools. The models are very different from each other on some of the nine components; however, one model is not necessarily better than another. Rather, administrators and teachers need to decide which components are important and then select the model that provides a consistent emphasis. To date, these models have been promulgated as packages; in the future, more research and practice is needed on defining and investigating their essential features.

REFERENCES

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Berk, R. (1984). *A guide to criterion-referenced test construction*. Baltimore, MD: Johns Hopkins University Press.

Berk, R. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research, 56*, 137-172.

Blankenship, C. (1985). Using curriculum-based assessment data to make instructional decisions. *Exceptional Children, 52*(3), 233-238.

Bursuck, W. D., & Lessen, E. (1985). Curriculum-based assessment: Increasing the participation of regular educators in mainstreaming. *ICEC Quarterly, 34*, 25-26.

Bursuck, W. D., & Lessen, E. (1987). A classroom-based model for assessing students with learning disabilities. *Learning Disabilities Focus, 3*(1), 17-29.

Carey, L. (1988). *Measuring and evaluating school learning*. Boston: Allyn & Bacon.

Casey, A., Deno, S. L., Marston, D., & Skiba, R. (1988). Experimental teaching: Changing teacher beliefs about effective instructional practices. *Teacher Education and Special Education, 11*(3), 123-132.

Conoley, J. C., & Kramer, J. J. (Eds.). (1989). *The tenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Coulter, A. (1985). Implementing curriculum-based assessment: Considerations for pupil appraisal professionals. *Exceptional Children, 52*(3), 277-281.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219-232.

Deno, S. L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review, 15*(3), 358-374.

Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. Shinn (Ed.), *Curriculum-based measurement*. New York: Guilford Press, School Psychology Practitioner's Series.

Deno, S. L., Chiang, B., Tindal, G., & Blackburn, M. (1979). *Experimental analysis of program components: An approach to research in CSDCs* (Research Report No. 12). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

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., Lowry, L., Sindelar, P., & Jenkins, J. (1982). *The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study* (Research Report No. 87). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Deno, S. L., Marston, D., Shinn, M. R., & Tindal, G. (1983). Oral reading fluency: A simple datum for scaling reading disability. *Topics in Learning and Learning Disabilities, 2*(4), 53-59.

Deno, S. L., Marston, D., & Tindal, G. (1985/86). Direct and frequent curriculum-based measurement: An alternative for educational decision making. *Special Services in the Schools, 2*(2/3), 5-27.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.

Deno, S. L., Mirkin, P., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*(1), 36-45.

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. (1980). *Relationships among simple measures of reading and performance on standardized achievement tests* (Research Report No. 20). Minneapolis, MN:

University of Minnesota, Institute for Research on Learning Disabilities.

Deno, S. L., Mirkin, P. K., Lowry, L., & Kuehnle, K. (1980). *Relationships among simple measures of spelling and performance on standardized achievement tests* (Research Report No. 21). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Deno, S. L., Mirkin, P. K., & Marston, D. (1980). *Relationships among simple measures of written expression and performance on standardized achievement tests* (Research Report No. 22). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Deno, S. L., Mirkin, P., & Shinn, M. (1979). *Behavioral perspectives on the assessment of learning disabled children* (Monograph No. 12). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Ebel, R. (1982). Proposed solutions to two problems of test construction. *Journal of Educational Measurement, 19*, 267-278.

Ebel, R., & Frisbie, D. (1986). *Essentials of educational measurement.* Englewood Cliffs, NJ: Prentice-Hall.

Enell, N. C., & Barrick, S. W. (1983). *An examination of the relative efficiency and usefulness of computer-assisted individualized education programs.* Carmichael, CA: San Juan Unified School District. (ERIC Document Reproduction Service No. ED 236-861)

Freeman, D. J., Kuhs, T. M., Porter, A. C., Floden, R. E., Schmidt, W. H., & Schwille, J. R. (1983). Do textbooks and tests define a national curriculum in elementary school mathematics? *The Elementary School Journal, 83*(5), 501-513.

Fuchs, L. S. (1986). Monitoring progress among mildly handicapped pupils: Review of current practice and research. *Remedial and Special Education, 7*, 5-12.

Fuchs, L. S. (1989). Evaluating solutions: Monitoring progress and revising intervention plans. In M. Shinn (Ed.), *Curriculum-based measurement.* New York: Guilford Press, School Practitioner Series.

Fuchs, L. S., Deno, S. L., & Marston, D. (1983). Improving the reliability of curriculum-based measures of academic skills for psychoeducational decision making. *Diagnostique, 8*(3), 135-149.

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

Fuchs, L. S., Fuchs, D., & Deno, S. L. (1982). Reliability and validity of curriculum-based informal reading inventories. *Reading Research Quarterly, 18*(1), 6-26.7.

Fuchs, L. S., Fuchs, D., & Deno, S. L. (1984). Sampling passages and student behavior for IRIs: A reply to Dufflemeyer. *Reading Research Quarterly, 19*(2), 248-249.

Fuchs, L. S., Fuchs, D., & Deno, S. L. (1985). The importance of goal ambitiousness and goal mastery to student achievement. *Exceptional Children, 52,* 63-71.

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

Fuchs, L. S., & Shinn, M. (1989). Writing CBM IEP objectives. In M. Shinn (Ed.), *Curriculum-based measurement.* New York: Guilford Press, School Psychology Practitioner's Series.

Fuchs, L. S., Tindal, G., & Deno, S. L. (1981). *Effects of varying item domain and sample duration on technical characteristics of daily measure in reading* (Research Report No. 48). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Germann, G. (1986a). *Continuous assessment program—reading/spelling.* Cambridge, MN: Performance Monitoring Systems.

Germann, G. (1986b). *Continuous assessment program—math.* Cambridge, MN: Performance Monitoring Systems.

Germann, G., & Tindal, G. (1985). Applications of direct and repeated measurement using curriculum based assessment. *Exceptional Children, 51*(2), 110-121.

Gickling, E., & Armstrong, D. L. (1978). Levels of instructional difficulty as related to on-task completion and comprehension. *Journal of Learning Disabilities, 11,* 559-566.

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

Gickling, E. E., & Thompson, V. P. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52*(3), 205-218.

Glass, G. (1980). When educators set standards. In E. L. Baker & E. S. Quellmalz (Eds.), *Educational testing and evaluation* (pp. 185-193). Beverly Hills, CA: Sage Publications.

Good, R. H. I., & Salvia, J. (1988). Curriculum bias in published, norm-referenced reading tests: Demonstrable effects. *School Psychology Review, 17*(1), 51-60.

Hambleton, R., & Swaminathan, H. (1978). Criterion-referenced testing and measurement: A review of technical issues and developments. *Review of Educational Research, 48,* 1-47.

Hargis, C. H. (1987). *Curriculum-based assessment: A primer.* Springfield, IL: Charles C. Thomas.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hopkins, C., & Antes, R. (1978). *Classroom measurement and evaluation.* Itasca, IL: F. E. Peacock.

Howell, K., & Morehead, M. (1987). *Curriculum-based evaluation for special and remedial education.* Columbus, OH: Charles Merrill.

Idol, L., Nevin, A., & Paolucci-Whitcomb, P. (1986). *Models of curriculum-based assessment.* Rockville, MD: Aspen.

Jenkins, J. R., & Pany, D. (1978). Curriculum biases in reading achievement tests. *Journal of Reading Behavior, 10*(4), 345-357.

Jenkins, M. W. (1987). Effect of an individualized education program (IEP) writer on time savings and quality. *Journal of Special Education Technology, 8*(3), 55-66.

Johnson, M., Kress, R., & Pikulski, J. (1987). *Informal reading inventories* (2nd ed.). Newark, DE: International Reading Association.

Krauss, J. (1988). *Investigating written retells as a measure of reading comprehension.* (Unpublished Master's Thesis), University of Oregon, Eugene, OR.

Leinhardt, G., & Seewald, A. M. (1981). Overlap: What's tested, what's taught? *Journal of Educational Measurement, 18*(2), 85-96.

Lindsley, O. (1964). Direct measurement and prosthesis of retarded behavior. *Journal of Education, 147,* 62.

Livingston, S., & Zieky, M. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests.* Princeton, NJ: Educational Testing Service.

Marston, D. (1988a). The effectiveness of special education: A time series analysis of reading performance in regular and special education. *The Journal of Special Education, 21*(4), 13-26.

Marston, D. (1988b). Measuring progress on IEPs: A comparison of graphing approaches. *Exceptional Children, 55*(1), 38-44.

Marston, D., & Deno, S. L. (1981). *The reliability of simple, direct measures of written expression* (Research Report No. 50). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Marston, D., & Deno, S. L. (1982). *Measuring academic progress of students with learning difficulties: A comparision of the semi-logarithmic chart and equal interval graph paper* (Research Report No. 101). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Marston, D., Fuchs, L. S., & Deno, S. L. (1985). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-based measures. *Diagnostique, 11,* 77-90.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52*(3), 266-276.

McKenna, M., & Robinson, R. (1980). *An introduction to the cloze procedure.* Newark, DE: International Reading Association.

Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher,* 9-19.

Mirkin, P. K., & Deno, S. L. (1979). *Formative evaluation in the classroom: An approach to improving instruction* (Research Report No. 10). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Mirkin, P. K., Deno, S., Fuchs, L. S., Wesson, C., Tindal, G., Marston, D., & Kuehnle, K. (1981). *Procedures to develop and monitor progress on I.E.P. goals* (training manual). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. (1980). *Formative evaluation: Continued development of data utilization systems* (Research Report No. 23). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Mirkin, P. K., Fuchs, L. S., & Deno, S. L. (1982). *Considerations for designing a continuous evaluation system: An integrative review* (Monograph No. 20). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Mitchell, J. (1985). *The ninth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

Nunnally, J. (1967). *Psychometric theory.* New York: McGraw-Hill.

Parker, R., Hasbrouck, J., & Tindal, G. (1989). *Combining informal teacher judgment and objective test scores to make cross-classroom reading group placements* (Research Report No. 4). Eugene, OR: University of Oregon Resource Consultant Training Program.

Parsonson, B.S., & Baer, D.M. (1978). The analysis and presentation of graphic data. In T. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (pp. 101-166). New York: Academic Press.

Popham, J. (1972). *Developing IOX objectives-based tests: Procedural guidelines.* (Technical Paper No. 8). Los Angeles: Instructional Objectives Exchange.

Popham, J. (1978). As always, provocative. *Journal of Educational Measurement, 15,* 297-300.

Popham, W.J. (1984). Specifying the domain of content or behaviors. In R.A. Berk (Ed.), *A guide to criterion-referenced test construction.* (pp. 29-48). Baltimore: Johns Hopkins University Press.

Posavac, E., & Carey, R. (1989). *Program evaluation: Methods and case studies.* Englewood Cliffs, NJ: Prentice-Hall.

Roid, G., & Haladyna, T. (1982). *A technology for test item writing.* New York: Academic Press.

Rosenfield, S. (1987). *Instructional consultation.* Hillsdale, NJ: Lawrence Erlbaum.

Rosenfield, S., & Rubinson, F. (1985). Introducing curriculum-based assessment through consultation. *Exceptional Children, 52*(3), 282-287.

Ryan, L. B., & Rucker, C. N. (1986). Computerized vs. noncomputerized individualized education programs: Teachers' attitudes, time, and cost. *Journal of Special Education Technology, 8*(1), 5-12.

Salvia, J., & Ysseldyke, J. (1988). *Assessment in special and remedial education.* Boston, MA: Houghton Mifflin.

Scruggs, T. E., Mastropieri, M., & Castro, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education, 8*(2), 24-33.

Shapiro, E. (1989). *Academic skills problems: Direct assessment and intervention.* New York: Guilford Press, School Practitioner Series.

Shapiro, E. S., & Derr, T. F. (1987). An examination of overlap between reading curricula and standardized achievement tests. *Journal of Special Education, 21*(2), 59-67.

Scheffler, I. (1965). Philosophical models of teaching. *Harvard Educational Review, 35*, 131-143.

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

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

Shinn, M. R., Gleason, M., & Tindal, G. (1989). Varying the difficulty of testing materials: Implications for curriculum-based measurement. *Journal of Special Education, 23*(2), 223-233.

Shinn, M. R., Tindal, G., & Stein, S. (1988). Curriculum-based assessment and the identification of mildly handicapped students: A research review. *Professional School Psychology, 3*(1), 69-85.

Skiba, R., & Deno, S. L. (1983). *A correlational analysis of the statistical properties of time-series data and their relationship to student achievement in resource classrooms* (Research Report No. 138). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Skiba, R., Deno, S. L., Marston, D., & Casey, A. (1989). Influence of trend estimation and subject familiarity on practitioners' judgments of intervention effectiveness. *Journal of Special Education, 22*(4), 433-446.

Skiba, R., Marston, D., Wesson, C., Sevcik, B., & Deno, S. L. (1983). *Characteristics of the time-series data collected through curriculum-based reading measurement* (Research Report No. 125). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Tawney, J., & Gast, D. (1984). *Single subject research in special education.* Columbus, OH: Charles Merrill.

Tindal, G. (1983). *Factors influencing the agreement between visual and statistical analyses of time-series data.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis, MN.

Tindal, G. (1984). *Procedures for establishing IEP goals and objectives using curriculum-based measurement.* Unpublished manuscript. Mpls, MN: University of Minnesota.

Tindal, G. (1987). Graphing performance. *Teaching Exceptional Children,* 20(1), 44-46.

Tindal, G. (1988). Curriculum-based assessment. In J. Graden, J. Zins, & M. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 111-136). Kent, OH: National Association of School Psychologists.

Tindal, G. (1989). Evaluating the effectiveness of educational programs at the systems level using curriculum-based measurement. In M. Shinn (Ed.), *Curriculum-based measurement.* New York: Guilford Press, School Practitioner Series.

Tindal, G., & Deno, S. L. (1981). *Daily measurement of reading: Effects of varying the size of the item pool* (Research Report No. 55). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Tindal, G., & Deno, S. L. (1983). *Factors influencing the agreement between visual and statistical analyses of time-series data* (Research Report No. 118). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Tindal, G., Deno, S. L., & Ysseldyke, J. (1983). *Visual analysis of time series data: Factors of influence and levels of reliability* (Research Report No. 112). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Tindal, G., Fuchs, L. S., Christenson, S., Mirkin, P., & Deno, S. L. (1981). *The relationship between student achievement and teacher assessment of short- or long-term goals* (Research Report No. 61). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Tindal, G., Fuchs, L. S., Fuchs, D., Shinn, M., Deno, S. L., & Germann, G. (1985). Empirical validation of criterion referenced tests. *Journal of Educational Research, 78*, 203-209.

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

Tindal, G., & Marston, D. (1990). *Classroom-based assessment: Evaluating instructional programs.* Columbus, OH: Charles Merrill.

Tindal, G., Marston, D., Deno, S. L., & Germann, G. (1982). *Curriculum differences in direct repeated measures of reading* (Research Report No. 93). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Tindal, G., & Parker, R. (1989). Assessment of written expression for students in compensatory and special education programs. *The Journal of Special Education, 23*(2), 169-184.

Tindal, G., & Parker, R. (1989). Development of written retell as a curriculum-based measure in secondary programs. *School Psychology Review, 18*(3), 328-343.

Tindal, G., Shinn, M., & Germann, G. (1987). The effects of different metrics on the interpretation of change in program evaluation. *Remedial and Special Education, 8*(5), 19-28.

Tucker, J. (1985). Curriculum-based assessment: An introduction. *Exceptional Children, 52*(3), 199-204.

Videen, J., Deno, S. L., & Marston, D. (1982). *Correct word sequences: A valid indicator of proficiency in written expression* (Research Report No. 84). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities.

Wesson, C. (1989). An efficient technique for establishing reading groups. *Reading Teacher, 42*(7), 466-469.

White, O. (1987). Some comments concerning "The qualitative synthesis of single subject research." *Remedial and Special Education, 8*(2), 34-39.

White, O., & Haring, N. (1980). *Exceptional Teaching.* Columbus, OH: Charles Merrill.

Witt, J., Elliott, S., Gresham, F., & Kramer, J. (1988). *Assessment of special children: Tests and the problem solving process.* Glenview, IL: Scott, Foresman.

Ysseldyke, J., & Thurlow, M. (1984). Assessment practices in special education: Adequacy and appropriateness. *Educational Psychologist, 19*, 123-137.

3

Enhancing Instructional Programming and Student Achievement with Curriculum-Based Measurement

Lynn S. Fuchs

*George Peabody College
of Vanderbilt University*

Curriculum-based measurement (CBM) is a form of curriculum-based assessment. As such, CBM has three features in common with all curriculum-based assessment approaches (Tucker, 1987): Test stimuli are drawn from the student's curriculum; assessment is ongoing and repeated across time; and assessment data are used to formulate instructional decisions.

Despite these similarities to other forms of curriculum-based assessment, CBM is distinctive because of two important features: It measures student proficiency across the annual curriculum and relies on standardized, prescriptive measurement methods (Fuchs & Deno, in press). The first purpose of this chapter is to explain these two features of CBM, by contrasting the CBM model to the predominant, mastery measurement form of curriculum-based assessment.

The second objective of this paper is to demonstrate how CBM databases can be used to help formulate instructional decisions. Within this context, research investigating the efficacy of each instructional use is reviewed.

THE CURRICULUM-BASED MEASUREMENT MODEL

As indicated above, two important features of curriculum-based measurement (CBM) are (a) its focus on measuring student proficiency across the annual curriculum and (b) its use of a standardized, prescriptive measurement methodology, with demonstrated psychometric acceptability. To explain each of these features, I contrast CBM to the more common, predominant form of curriculum-based assessment known as mastery measurement. Within this section, I first explain and provide an example of mastery measurement. Then, I explain CBM and provide an example. Finally, the salient differences between mastery measurement and CBM are explored.

Mastery Measurement

Mastery measurement is the most common form of curriculum-based assessment (see Shinn, Rosenfield, & Knutson, 1989 for discussion of different types of curriculum-based assessment). Mastery measurement describes student mastery of a series of short-term instructional objectives or instructional levels (see Blankenship, 1985 and Gickling & Thompson, 1985 for explanation of these forms of mastery measurement). So, for example, let us say that Mrs. P. wants Dolly to master the fourth-grade computation curriculum. That is, by June Mrs. P. wants Dolly to compute accurately all problem types encompassed within the fourth-grade curriculum. In designing a mastery measurement system, Mrs. P. would begin by completing two large tasks. She would (a) determine a sensible instructional sequence for the fourth-grade computation curriculum and (b) design a criterion-referenced testing procedure to match each step in that instructional sequence.

Let us say, for example, that after careful inspection of the fourth-grade computation curriculum, Mrs. P. identified the skills listed in Table 1. These are the universe of problem types incorporated within her fourth-grade curriculum. She further determined that a logical sequence of skills for instruction were the following: multidigit addition with regrouping, multidigit subtraction with regrouping, multiplication facts (factors to 9), division facts (divisors 6-9), multiplying two 2-digit numbers without regrouping, multiplying 1- or 2-digit numbers with regrouping, dividing 3- by 1-digit numbers without remainders, dividing 2- or 3- by 1-digit numbers with remainders, adding and subtracting mixed decimals to hundredths, and adding and subtracting simple or mixed fractions without regrouping.

Table 1
Fourth Grade Curriculum

Sequence	Skill	Proportion
1	Multidigit addition with regrouping	12%
2	Multidigit subtraction with regrouping	4%
3	Multiplication facts, factors to 9	24%
4	Division facts, divisors 6-9	16%
5	Multiplying two 2-digit numbers, no regrouping	4%
6	Multiplying 1- or 2-digit numbers, with regrouping	12%
7	Dividing 3- by 1-digit numbers, no remainder	4%
8	Dividing 2- or 3- by 1-digit numbers, with remainder	4%
9	Adding and subtracting mixed decimals to hundredths	8%
10	Adding and subtracting simple or mixed fractions, no regrouping	12%

Having established the instructional sequence, Mrs. P.'s second major task in establishing a mastery measurement system would be to design a criterion-referenced testing procedure for each step in her instructional hierarchy. By definition, Mrs. P. would begin by measuring the first skill in the sequence, multidigit addition with regrouping. She decides on a criterion-referenced assessment procedure that involves preparing 25 comparable tests, each containing 10 problems that feature multidigit addition with regrouping. To maintain a moderate degree of comparability in the difficulty of the items on this "multidigit addition" test, Mrs. P. decides that all problems will present 3- or 4-digit numerals. The criterion-referenced testing procedure will involve presenting the test, along with directions, allowing 3 minutes for writing answers, and scoring performance in terms of the number of correct problems written in 3 minutes. Mrs. P. defines mastery as eight correct problems in 3 minutes on 3 consecutive days. (In a similar way, Mrs. P. would design a criterion-referenced testing procedure to assess mastery of each problem type listed in Table 1.)

Having ordered the skills embedded in the curriculum and having designed a criterion-referenced testing procedure for each skill in the

instructional sequence, Mrs. P. would teach multidigit addition with regrouping and test Dolly's proficiency on this problem type on a regular basis. When Dolly achieves mastery of multidigit addition with regrouping, Mrs. P. simultaneously would shift instruction and measurement to the next teaching step: multidigit subtraction requiring regrouping. A mastery measurement graph, illustrating Mrs. P.'s measurement system for Dolly, is shown in Figure 1.
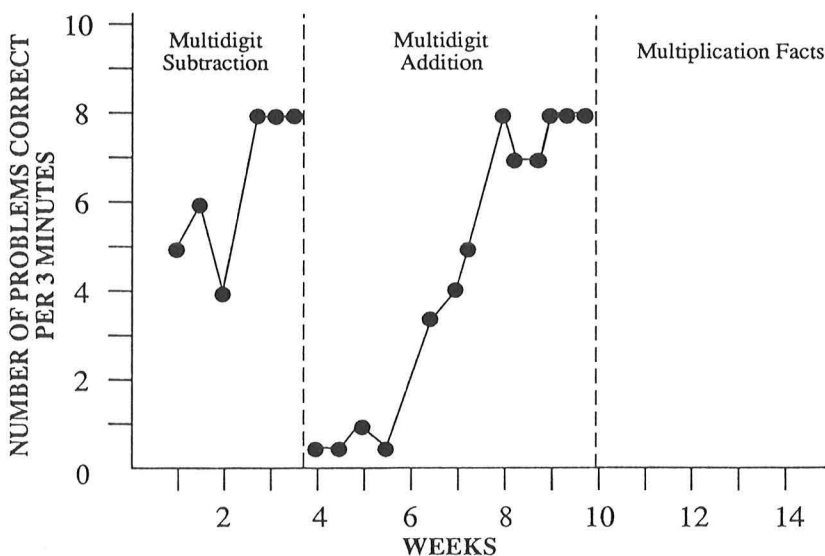
Figure 1. Example of a mastery measurement graph.

As depicted in this figure, it took 3 1/2 weeks of instructional time before Dolly demonstrated mastery of multidigit addition with regrouping. Then, when mastery of multidigit addition was achieved, Mrs. P. shifted instruction and measurement to the second step of the instructional hierarchy: multidigit subtraction. Approximately 6 weeks later, when mastery of multidigit subtraction was demonstrated, Mrs. P. began instruction on the third skill of the hierarchy, multiplication of basic facts (factors to 9). Consequently, measurement would be conducted on the criterion-referenced testing approach Mrs. P. designed to assess proficiency on multiplication facts (factors to 9).

Curriculum-Based Measurement

As distinguished from the predominant form of curriculum-based assessment, (i.e., mastery measurement), two important characteristics of curriculum-based measurement (CBM) are (a) assessment of proficiency on skills that represent the entire, year-long curriculum and (b) reliance on standardized, prescriptive measurement methods. To clarify, let me return to the example of Mrs. P. and Dolly.

In this case, Mrs. P. maintained her goal for Dolly (i.e., proficiency on the fourth-grade computation curriculum), but she decided to rely on CBM rather than on mastery measurement. Instead of sequencing the fourth-grade computation curriculum and formulating a criterion-referenced testing procedure for each step in the instructional sequence, Mrs. P. would complete the following process.

Sheet #30
Password: JAR
Name : _____   Date : _____

A	B	C	D	E
$\begin{array}{r} 747 \\ \times\ 3 \end{array}$	$\begin{array}{r} 5406 \\ 4721 \\ +\ \ \ 75 \end{array}$	$\begin{array}{r} 1 \\ \times 1 \end{array}$	$\begin{array}{r} 2666 \\ -\ 2647 \end{array}$	$9\overline{)27}$
F	G	H	I	J
$5\overline{)800}$	$2\frac{1}{3} + 6 =$	$\begin{array}{r} 0 \\ \times\ 4 \end{array}$	$2\frac{2}{5} - 1 =$	$\begin{array}{r} 5 \\ \times\ 4 \end{array}$
K	L	M	N	O
$\begin{array}{r} 1670 \\ 4121 \\ +\ \ \ 86 \end{array}$	$7\overline{)21}$	$4\overline{)962}$	$7\overline{)28}$	$\begin{array}{r} 26.8 \\ +\ 13.35 \end{array}$
P	Q	R	S	T
$\begin{array}{r} 51 \\ \times 91 \end{array}$	$\begin{array}{r} 0 \\ \times 3 \end{array}$	$\begin{array}{r} 4 \\ \times 4 \end{array}$	$\frac{2}{3} - \frac{1}{3} =$	$\begin{array}{r} 33 \\ \times 22 \end{array}$
U	V	W	X	Y
$7\overline{)7}$	$\begin{array}{r} 16.42 \\ -\ 3.8 \end{array}$	$\begin{array}{r} 702 \\ \times\ 7 \end{array}$	$\begin{array}{r} 46943 \\ +\ 80950 \end{array}$	$\begin{array}{r} 5 \\ \times 8 \end{array}$

Figure 2. Example of CBM computation test.

She would list (a) the problems that constitute the fourth-grade computation curriculum and (b) the proportion of problem types that accurately represent the curriculum. For the statewide Tennessee "Basic Skills First" fourth-grade curriculum, these problem types and corresponding proportions are shown in Table. 1. This pool of problem types is the domain that Mrs. P. wants Dolly to master by June; it is Dolly's annual, year-long curriculum. Then, according to CBM methodology (Fuchs, Fuchs, & Hamlett, 1989a), Mrs. P. would use randomly generated numerals to create a series of alternate test forms. Each test would comprise 25 problems that represent the type and proportion of problems constituting the fourth-grade curriculum. One alternate form of the fourth-grade computation test is shown in Figure 2. To accomplish the test-construction process, Mrs. P. could use a computer program (Fuchs, Hamlett, & Fuchs, 1990). With this program, Mrs. P. would specify the problem types and proportions to the computer; the computer would generate the alternate forms. Then, according to standard CBM methodology (Fuchs, Fuchs, & Hamlett, 1989a), Mrs. P. would administer and score each CBM test in the following way. She would present a test and a standard set of directions to the student, and allow Dolly 3 minutes to complete as much of the test as possible. Mrs. P. would score performance in terms of the number of digits Dolly wrote correctly in 3 minutes.

Each math test samples the year-long domain in the same way; each test is an alternate form that represents the fourth-grade curriculum. As shown in Figure 2, the CBM test samples computation behaviors *across* the skills representing the fourth-grade curriculum (these skills are listed in Table 1). During the first part of the school year (i.e., in October), Dolly has poor mastery of the fourth-grade curriculum, and her scores are low on the CBM test (i.e., 18 digits correct; see scores shown in Figure 3). The total number of correct digits score on the CBM test is a performance indicator of Dolly's overall proficiency in the fourth-grade computation curriculum. The score does not communicate *which* skills in the curriculum have and have not been mastered; rather, it *indicates* that few skills are mastered. The teacher can, however, determine Dolly's specific skill profile using the CBM database. The practitioner can analyze Dolly's performance on the specific items on the CBM tests, which sample across the fourth-grade curricular skills, to determine which skills currently are mastered. When the teacher conducts such an item analysis on the CBM tests, he/she corroborates the lack of proficiency indicated by the score of 18. As shown in Table 2, which displays the profile of skills achieved at three points in time across the year, when the practitioner analyzes the responses on the

items of the test, the performance indicator of 18 is associated with no mastered skills and only several partially mastered skills.

As the year progresses and instruction continues, Dolly's CBM scores increase gradually. By February, Dolly has earned scores of 45 digits correct (see Figure 3). When we analyze the responses on the CBM tests, we see that this increased score of 45 digits is associated with three mastered skills, five partially mastered, and only two nonmastered skills in the fourth-grade curriculum. Then, as time passes and additional instruction occurs, Dolly gains proficiency on the fourth-grade curriculum; her performance indicator continues to increase to 55 by April (see Figure 3), and the profile of fourth-grade skills mastered concurrently improves (see Table 2).

Within CBM, the performance indicators are presented in graphic form. For example, the graph in Figure 3 shows Dolly's scores on the CBM tests across time. As the year progresses, Dolly's scores increase. The slope of Dolly's scores across time represents Dolly's overall learning rate in the fourth-grade curriculum. As the performance indicator (or CBM score) increases, Mrs. P. knows that Dolly's overall proficiency in the fourth-grade curriculum has increased, and she has confidence that Dolly's mastery of specific fourth-grade skills also is improving.
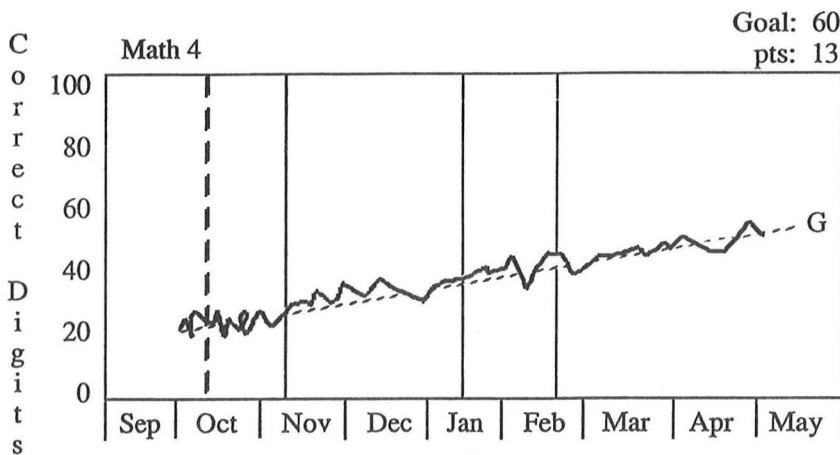
Figure 3.

Table 2

Skills Profile at Three Points in Time

Date	Mastered	Partially Mastered	Nonmastered	Not Attempted
October		Multidigit addition, regrouping Multiplication facts Multiplication 1- or 2-digit, regrouping	Multidigit subtraction, regrouping Multiplication, no regrouping Division facts Dividing 3- by 1-digit, no remainder	Dividing 2- or 3- by 1-digit, remainder Adding/subtracting simple/mixed fractions, no regrouping Adding/subtracting mixed decimals to hundredths
February	Multidigit addition, regrouping Multiplication facts Multidigit subtraction, regrouping	Multiplication 1- or 2- digit, regrouping Division facts Dividing 3- by 1-digit, no remainder Dividing 2- or 3- by 1-digit, remainder Adding/subtracting mixed decimals to hundredths	Multiplication, no regrouping Adding/subtracting simple/mixed fractions, no regrouping	
April	Multidigit addition, regrouping Multiplication facts Multidigit subtraction, regrouping Division facts Dividing 2- or 3-digit by 1-digit, remainder Adding/subtracting mixed decimals to hundredths	Multiplication, no regrouping Multiplication 1- or 2-digit, regrouping Dividing 3- by 1-digit, no remainder	Adding/subtracting simple/mixed fractions, no regrouping	

Important Distinctions Between Mastery Measurement and CBM

Five important distinctions exist between mastery measurement and CBM. These salient differences are (a) the scope of skills upon which measurement is focused, (b) the extent to which generalization and maintenance are assessed, (c) the degree of constancy in measurement across time, (d) the reliance of the measurement on instructional hierarchies, and (e) the methods by which measurement methods are developed. An explanation of each of these differences follows.

Scope of skills for measurement. Mastery measurement and CBM are essentially different because of the scope of skills encompassed within these two forms of measurement. Specifically, mastery measurement is relatively narrow; it focuses measurement on single skills (or small clusters of skills) at a time. By contrast, CBM is relatively broad; it focuses measurement on a large domain of skills, representing the curriculum to be mastered over the course of a school year.

Mastery measurement focuses instruction and measurement on a series of short-term instructional objectives; therefore, instruction and measurement are linked together. An advantage of this linking is that the assessment data should be highly sensitive, or responsive, to instructional effects. This indicates strong instructional validity (Yalow & Popham, 1983). Nevertheless, a potential disadvantage of a close connection between measurement and instruction is that the measurement framework is restricted. Scores may reflect the student's skill in computing only in the narrow framework within which testing occurs (i.e., when all problems require use of the same multidigit-regrouping addition algorithm). So, the content validity, reflecting the extent to which the measurement mirrors the domain—computing problems in natural or mixed presentation—may be reduced. Also, the relation between progress through an instructional sequence and socially important outcomes, such as standardized, commercial achievement test performance, is uncertain.

In contrast, CBM focuses on the long-term goal. That is, rather than measuring student mastery on a series of changing instructional objectives, CBM focuses measurement on the relatively broad, annual curriculum. The disadvantage associated with such a broad focus is the loss of potential instructional validity. Compared to mastery measurement, where the teacher tests performance on the immediate instructional objective, CBM samples content across the year-long curriculum. Consequently, CBM may be less sensitive than mastery measurement to student change as a result of current instruction (Fuchs

& Deno, in press). However, compared to traditional measurement, where performance samples behavior across both grade levels and curricula at one moment in time, CBM provides information that (a) is sensitive to instructional effects (Marston, Fuchs, & Deno, 1985) and (b) can be used to improve instructional decision making (Fuchs & Fuchs, 1990).

Also, as can be anticipated in light of the foregoing discussion, CBM's focus on long-term goal measurement offers certain advantages over mastery measurement. Because CBM describes student performance in terms of proficiency on the annual curriculum, both its content and criterion validity are stronger than mastery measurement (Fuchs & Fuchs, 1986).

Retention and generalization of skills. A second key distinction between mastery measurement and CBM is the extent to which the measurement assesses retention and generalization of skills. With mastery measurement's close connection between testing and instruction, mastery measurement does not automatically assess retention and generalization of skills. When Dolly demonstrates mastery of multidigit addition with regrouping (and when measurement and instruction simultaneously shift to subtraction with regrouping), we have no automatic index of the extent to which Dolly retains mastery of multidigit addition. Conversely, while Mrs. P. focuses instruction and testing on multidigit addition, we have no indication of the extent to which Dolly may generalize her increasing skill in multidigit addition to other dimensions of the curriculum. For example, as Dolly gains mastery of multidigit addition with whole numbers, she may acquire skill in mixed addition of decimals to the hundredths place. Yet, a mastery measurement system will not index this generalization. As this illustrates and as Goodstein (1982) has described, closely linking the instructional format to assessment (or narrowly defining the content-x-format domain of criterion-referenced/mastery measurement) may create problems, including the failure to index retention and generalization learning events.   ⅄

In contrast to mastery measurement, CBM offers the advantage of automatically assessing retention and generalization of skills. As Dolly improves her skill in multidigit addition with regrouping, the CBM performance indicator should increase, because Dolly's increased proficiency allows her to compute the multidigit addition problems with regrouping (and therefore more digits) correctly on the CBM tests. However, if Dolly fails to retain mastery of multidigit addition with regrouping when multidigit subtraction with regrouping instruction begins, Dolly's CBM score should decrease. This would occur because

Dolly no longer would compute the multidigit addition with regrouping problems on the CBM tests correctly. Therefore, CBM is sensitive to retention because it samples skills *across* the annual curriculum.

Conversely, if Dolly generalizes learning to new skills when multidigit addition with regrouping instruction occurs, Dolly's performance indicators should increase, because opportunities for computing untaught problem types are provided on the CBM tests. In this way, CBM indexes generalization. This sensitivity of measurement to retention and generalization learning may be critical when CBM is used to monitor the development of basic skills for handicapped populations. These low-achieving pupils frequently have poorly developed strategies for maintaining and transferring skills (Anderson-Inman, Walker, & Purcell, 1984; White, 1984).

Constancy in measurement across time. A third difference between mastery measurement and CBM is the extent of constancy in measurement across time. Mastery measurement requires a shift in measurement each time a skill is mastered; CBM maintains a constant measurement focus across the year.

As shown in Figure 1, with the regular shifts in mastery measurement across time, we can determine an acquisition rate for multidigit addition with regrouping and we can estimate a separate learning curve for acquisition of multidigit subtraction with regrouping. However, it is impossible to summarize an overall learning rate *across* the different skills in the curriculum. This is because different skills, measured at different times during the school year, are not of equal difficulty and do not represent equal curriculum units. For example, research indicates that acquisition of subtraction skills is more difficult than mastery of addition skills. Consequently, one would not expect different skills (even seemingly analogous skills such as multidigit addition with regrouping and multidigit subtraction with regrouping) to be acquired in equivalent times. These unequal curriculum units, along with the shifts in measurement and the resultingly limited summaries of learning rate, appear to reduce the usefulness of mastery measurement.

With CBM, teachers may monitor students' basic skills development across a school year without any shifts in measurement. Because CBM tests sample across the entire year-long curriculum, test difficulty remains constant across the school year. As shown in Figure 3, the difficulty of the CBM tests Dolly took in November is comparable to the difficulty of the tests she took in March. It is Dolly's *proficiency*, not the test difficulty, that increases. However, with mastery measurement, the measurement domains and the difficulty of testing material continually change as the instructional content changes. CBM avoids

these shifts in measurement domains, and this constancy associated with CBM permits summaries of student learning rates across time. The CBM database can be used to compare the effectiveness of different instructional components introduced at different times during the year (see subsequent discussion).

Reliance on instructional hierarchies. Another key distinction between mastery measurement and CBM is the extent to which they rely on instructional hierarchies to determine measurement. In order to establish mastery measurement systems, teachers are required to specify instructional hierarchies that dictate the sequence for instruction and measurement. Most instructional hierarchies rely on "scope and sequence" charts (see Salvia & Hughes, 1990, for procedures for specifying instructional hierarchies within mastery measurement). Such charts tend to be long and detailed, and require teachers to group across skills (Salvia & Hughes, 1990). Additionally, scope and sequence charts typically are based on logical, rather than empirical, analyses of skills development. The appropriateness of logically determined sequences of instruction for students, especially handicapped pupils who do not progress along predictable developmental sequences, is unknown. Moreover, as demonstrated in the discussion that follows, when instructional hierarchies *determine* measurement, teachers cannot use assessment information to evaluate the effectiveness of alternative instructional approaches.

As opposed to mastery measurement, CBM does not require teachers to specify instructional hierarchies before measurement occurs. To set up a CBM system, a teacher identifies the annual domain on which he/she expects the student to be proficient by June. This offers certain advantages. First, the difficult task of compartmentalizing and ordering the curriculum is circumvented. This eliminates teacher effort, and avoids possible errors in specifying instructional chunks and sequences that eventually may prove troublesome to individual student growth.

Second, in sharp contrast to mastery measurement, CBM does not *determine* instruction. The structure of mastery measurement specifies the order in which instruction must proceed, and one cannot progress to subsequent skills until mastery of the current skills is demonstrated. Moreover, as illustrated in the work of Salvia and Hughes (1990), the mastery measurement framework also typically results in a skills-oriented approach to instruction, and the order in which skills are taught is determined by measurement. With mastery measurement, the independent variable (instruction) and the dependent variable (measurement) are tied together, with both simultaneously focused on

skills. With CBM, measurement (the dependent variable) is not tied to and determined by the current instructional focus or procedure (the independent variable); therefore, measurement and instruction are not confounded. Because of this, CBM offers the advantage of permitting teachers to experiment with contrasting instructional chunks, sequences, and/or procedures: Teachers use the CBM database as the dependent variable by which they evaluate the effectiveness of contrasting instructional strategies.

Development of tests. The fifth feature that differentiates mastery measurement and CBM is test development procedures. Mastery measurement relies primarily on the use of teacher-made criterion-referenced tests. Such teacher-made criterion-referenced tests have unknown technical characteristics. And the time-consuming and costly nature of reliability and validity studies makes it difficult, if not impossible, to investigate the psychometric characteristics of teacher-constructed measures. Additionally, even when teachers rely on commercial criterion-referenced tests for mastery measurement, psychometric characteristics are uncertain. Hambleton and Eignor (cited in Berk, 1982) evaluated 11 popular, commercially available criterion-referenced tests. They found that these tests could be characterized as follows:

-About half of the publishers included information about the qualifications of individuals who prepared the objectives on which the tests were based.

-Item representativeness could not be established because of the absence of domain specifications.

-For item analysis, there were two problems: Too little explanation was offered for the choice of particular item statistics and for the specifics of item statistics usage; and item statistics were used in test construction, thereby "biasing" the content validity of the test in unknown ways.

-Test score reliability was not handled well in most manuals.

-Inappropriate, or no, information relative to the stated uses of the test scores was offered.

-Rationales and procedures for setting cutoff scores were not offered, and evidence usually was not provided for the validity of cutoff scores (e.g., did examinees classified as masters typically perform better than those classified as nonmasters on some appropriate external criterion measure?).

-Factors affecting the validity of scores were not offered in any manuals.

-Few manuals introduced the notion of error in test scores or classifications of examinees to mastery states.

These findings, based on examination of criterion-referenced test manuals, are corroborated by empirical work. Tindal, Fuchs, Fuchs, Shinn, Deno, and Germann (1985) conducted reliability and validity studies on criterion-referenced tests associated with four popular basal reading series. Findings indicated variable reliability and validity coefficients, with many indices failing to reach acceptable levels. Consequently, commercial criterion-referenced tests frequently fail to provide information with documented reliability and validity.

In contrast to typical mastery measurement approaches, a comprehensive research program (Deno & Fuchs, 1987; Shinn, 1989) has investigated the psychometric characteristics of alternative methods for sampling test stimuli from curriculum, administering and scoring tests, and summarizing and evaluating scores in prescriptive ways. From this research, a standard CBM methodology has been formulated (Mirkin et al., 1984). Consequently, when teachers have determined the curriculum they expect students to master over the course of the school year, CBM prescribes methods for creating, administering, scoring, and using tests that result in reliable and valid descriptions of students' basic skills growth in reading, spelling, written expression, and computation. This standardized, prescriptive measurement within CBM, with documented reliability and validity, contrasts sharply with the unknown psychometric features of the teacher-made criterion-referenced tests used within mastery measurement.

USING CURRICULUM-BASED MEASUREMENT TO DEVELOP EFFECTIVE INSTRUCTIONAL PROGRAMS

Research supports three strategies for using curriculum-based measurement (CBM) to assist teachers in developing instructional programs. First, teachers can use CBM to monitor the appropriateness of the goals they set and to ensure the use of realistic, but ambitious, goals. Second, CBM can be used to determine the adequacy of student progress, to determine whether instructional programs require adjustment, and to compare the effectiveness of alternative programmatic components. Finally, CBM databases can be used to draw profiles of skill strengths and weaknesses, in order to assist teachers in determining the nature of effective programmatic modifications. In the following sections, each of these applications is described and the relevant research base is reviewed.

Using CBM to Monitor and Adjust Goals

Research substantiates the effectiveness of using goals to improve instructional outcomes. Summarizing across a variety of goal-writing procedures and research methods, Hartley and Davies (1976) found that teaching with goals enhances student achievement. McNeil (1967), for example, demonstrated that teachers who employed behavioral objectives produced better academic growth with their students and were judged to be more successful in applying learning principles, compared to a control group of teachers who did not use goals.

The relevant literature suggests that one way in which goals may mediate enhanced achievement outcomes is by structuring evaluation activities. A well-written goal defines the parameters of measurement: The goal specifies the anticipated observable performance that is desired, the conditions under which the behavior will be demonstrated, and the criteria against which to judge performance (Bloom, Hastings, & Madaus, 1971; Gagne, 1964; Mager, 1975). Adding this structure to the evaluation process may help teachers generate frequent, relevant student performance data. With ongoing feedback to practitioners and students, teachers can formulate more effective instructional programs (Jenkins, Deno, & Mirkin, 1979), and students can recognize their own successful learning strategies more readily (Bandura, 1982; Peckham & Roe, 1977; Rosswork, 1977).

CBM attempts to take advantage of potential benefits associated with the use of goals. Within CBM, the structure of the goal establishes key dimensions of the measurement/evaluation system. First, as the teacher selects the goal, she specifies the point within the curriculum where the student is expected to be proficient by year's end. This level becomes the measurement pool from which stimuli for testing are drawn. Second, when setting the goal, the teacher simultaneously indicates the performance criterion she is equating with "proficiency." This performance criterion creates the structure against which the adequacy of student progress is judged within CBM.

Let us say, for example, that Mrs. P. determines she wants a second student, Michael, to be proficient in Grade 3 of the computation curriculum by the end of the school year. Using CBM, Mrs. P. would measure Michael's performance on an alternate test, comprising 25 problems that represent the type and proportion of problems in the same way each time she tested Dolly's proficiency in the curriculum. Let us also say that Mrs. P. equates "proficiency" for Michael in this curriculum with a score of 20 digits correct by April 15. Using CBM, Mrs. P. would set up a monitoring graph to create a record of Michael's

progress and to evaluate the adequacy of Michael's growth. As shown in Figure 4 (top panel), this graph displays Michael's initial, or baseline, performance in the target Grade 3 curriculum (see dots that show scores of 5, 9, and 6); it shows the goal (see the "G" placed at the desired score

Wait: Not enough data for analysis.

UH-oh! Make a teaching change.
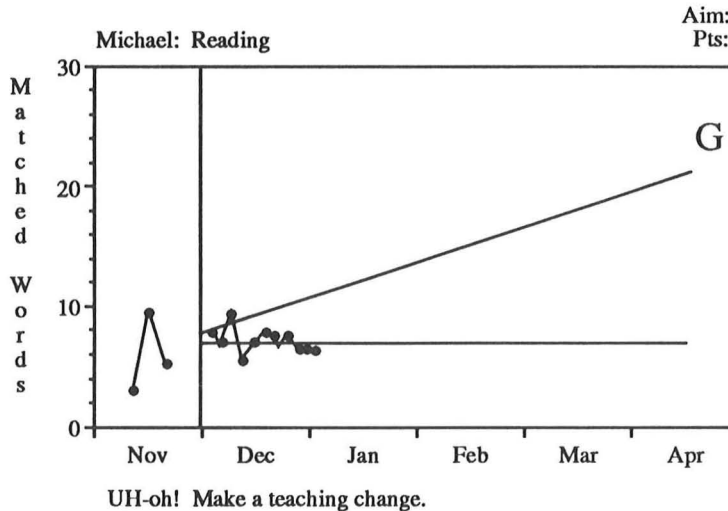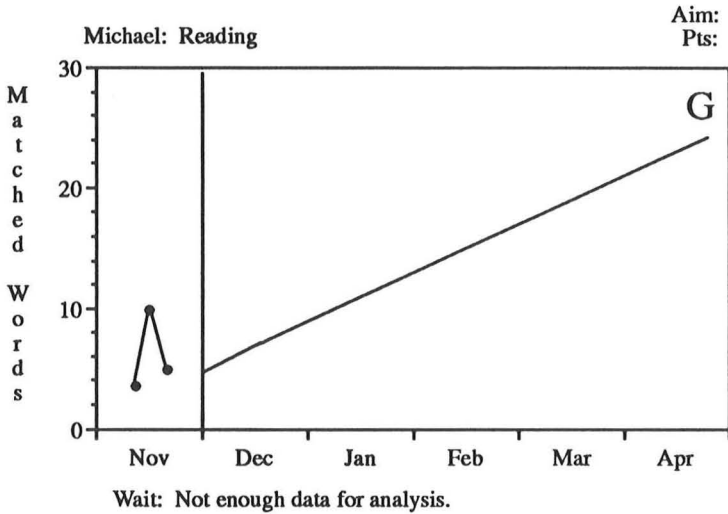
Figure 4. Example of CMB graphs. The top panel shows baseline, goal, and goal line; the bottom panel illustrates data for which the appropriate decision is for the teacher to make a teaching change.

of 20 on April 15); and it illustrates a "moving goal" (see the broken diagonal line) that indicates (a) the rate at which Michael will have to improve in order to attain the goal and (b) the target score on any given date.

Within typical CBM practice, the goal structures the evaluation process in the following way. When the student's actual rate of progress falls below the rate necessary for goal attainment, the rate of the student's progress and the effectiveness of the student's program are judged inadequate. In this case, CBM decision rules dictate that a teaching change is required. Figure 4 (bottom panel) shows an example of such a decision. Here the student's actual rate of progress, indicated by the solid diagonal line, is less steep than the desired rate of progress for goal attainment, indicated by the broken diagonal line. As illustrated, the decision in this case would be for the teacher to modify the instructional program in order to stimulate student progress.

As this discussion should make clear, the performance criterion specified in the goal becomes critical in the instructional decision-making process. Within the context of programming for handicapped or other low-achieving students, where the need for quality instructional programming is essential, the most critical potential problem associated with the performance-criterion-setting process may be the following: When teachers set goals that are unambitiously low, few if any recommendations for instructional improvements will be made.

Moreover, research indicates that unambitious goal setting within CBM relates to relatively poor student achievement. Fuchs, Fuchs, and Deno (1985) conducted a post-hoc analysis of a database in which each teacher, along with their four mildly to moderately handicapped pupils, had been assigned randomly to either a CBM or a control group condition for a 4-month study in the area of reading (Fuchs, Deno, & Mirkin, 1984). In this post-hoc study, student graphs were inspected after the completion of the CBM implementation. On the basis of inspecting graphs and looking at teachers' setting of goals and students' final performance levels, the 58 students in the CBM group were divided into three goal ambitiousness conditions: a highly ambitious goal group, a moderately ambitious goal group, and a low ambitious goal group. Students also were divided into two goal mastery conditions: those who had mastered and those who had not mastered their goals.

Three types of achievement outcomes were studied: (a) the Passage Reading Test, a measure that requires reading behavior similar to that required in the CBM tests; (b) the Stanford Diagnostic Reading Test, Structural Analysis subtest, a measure of decoding skills; and (c) the Stanford Diagnostic Reading Test, Reading Comprehension subtest. A

multivariate analysis of covariance, with appropriate follow-up analyses, indicated the following. The ambitiousness with which the goals were established was associated positively with student achievement. On two achievement measures, with pretreatment achievement levels statistically controlled, students for whom teachers set highly and moderately ambitious goals achieved better than students whose goals reflected relatively unambitious goals. On a third achievement measure, students with highly ambitious goals performed better than students for whom moderately ambitious and low goals were set. Furthermore, there were no effects associated with goal mastery. That is, students who met their goals and students who did not meet their goals achieved in comparable fashion. It was the level of goal *ambitiousness*, not goal *attainment*, that was associated with student achievement.

Based on these results, it appears that the selection of an appropriately ambitious, but realistic, performance criterion appears to be critical within CBM instructional decision making. Despite this importance, few satisfactory strategies for identifying appropriate performance criteria have been formulated. One potential solution to the goal-setting problem with CBM, referred to a dynamic goal setting, has been explored recently.

During the 1986-1987 academic year, Fuchs, Fuchs, and Hamlett (1989a) conducted a study designed to test the effectiveness of an innovative CBM goal-setting strategy, "dynamic" goal setting. In this study, participants were 30 special education teachers who taught self-contained and resource programs for students in Grades 2—9. Teachers selected two mildly handicapped students with IEP math goals. Then, teachers were assigned randomly to three treatment groups: dynamic goal CBM, static goal CBM, and control. The control teachers monitored student progress using conventional special education practice, including unit tests, correction of assignments, and unsystematic observation of student performance. The teachers in both CBM groups did the following. For 15 weeks, each teacher employed CBM to track their two pupils' progress toward math goals. The CBM system was rooted in the Tennessee Basic Skills First Math Program (BSF). The math computation objectives tested at each grade level within the BSF were listed. Teachers inspected these lists and determined an appropriate grade level on which to establish each student's goal. This level included the pool of math objectives the teacher hoped the student would master by year's end.

Using a standard measurement task, teachers were required to assess each pupil's math performance at least twice weekly, for 2 minutes, each time on a different probe representing the type and

proportion of problems from the BSF goal level they had selected. That is, if the teacher chose the third-grade level of the curriculum, the teacher was provided with 50 alternate test forms, each of which sampled the BSF third-grade computation objectives in the proportion tested on the BSF third-grade criterion-referenced end-of-year test. Each test could be conceptualized as a short form of the BSF third-grade computation test. Consequently, as teachers monitored pupil progress on these tests, they could estimate progress toward mastery of the corresponding level of the BSF end-of-year tests.

Each test was scored in terms of the number of correct digits written in 2 minutes. For half the students in each CBM group, scores were automatically collected using computers and saved to disk; for the other half, scores were collected by teachers and entered into a data-management software program by teachers. However, all testing procedures were completely analogous, and no outcome differences were associated with this administration factor (Fuchs, Hamlett, & Fuchs, 1987). Once each week, teachers used data-management software to review their students' assessment profiles. The software automatically graphed the scores, drew a goal, a goal line, and a regression line of best fit depicting the student's actual slope of improvement. Additionally, the software applied a set of decision rules. If the regression line was less steep than the goal line, the decision provided to the teacher read, "Uh-oh! Make a teaching change." When the regression line was steeper than the goal line, one of two possible decisions came up, depending on the teacher's experimental condition.

Within the *static goal* CBM group, when the student's actual rate of improvement exceeded the rate anticipated in the goal line, the decision read "OK! Collect more data." The data pattern suggested that the student's rate of progress was acceptable with respect to goal attainment, and that the corresponding instructional program looked effective. Thus, the message indicated that the teacher should keep the current instructional program intact and continue data collection. The teachers always were free to increase their goal, but they never were directed to do so. Figure 5 (top panel) shows a graph depicting satisfactory progress, and the message that would have been delivered within the static goal CBM condition.

Within the *dynamic goal* CBM group, when the student's actual rate of improvement exceeded the rate anticipated in the goal line, the decision read "OK! Raise the goal to X" (where X = the student's predicted performance at the end of the study, based on the student's current rate of progress). Again, the data pattern suggested that the

Michael: Math 3   Aim: 30
                                                             Pts: 9
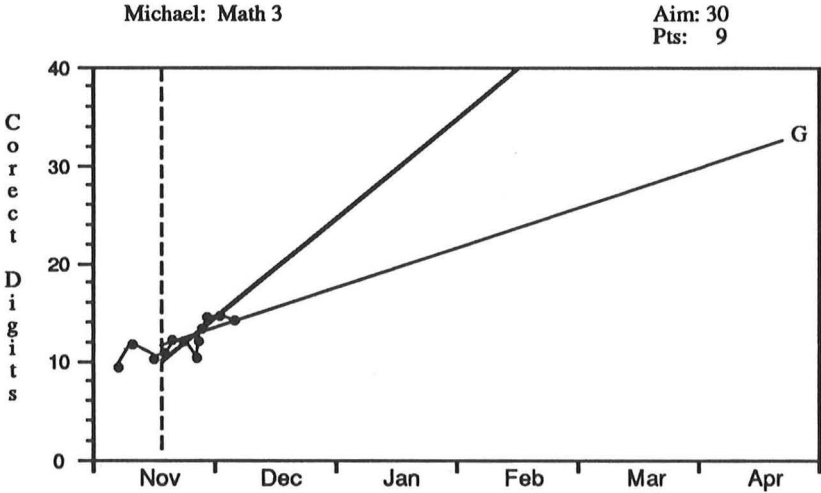
What should you do?

Keep collecting data.

Michael: Math 3   Aim: 30
                                                             Pts: 10
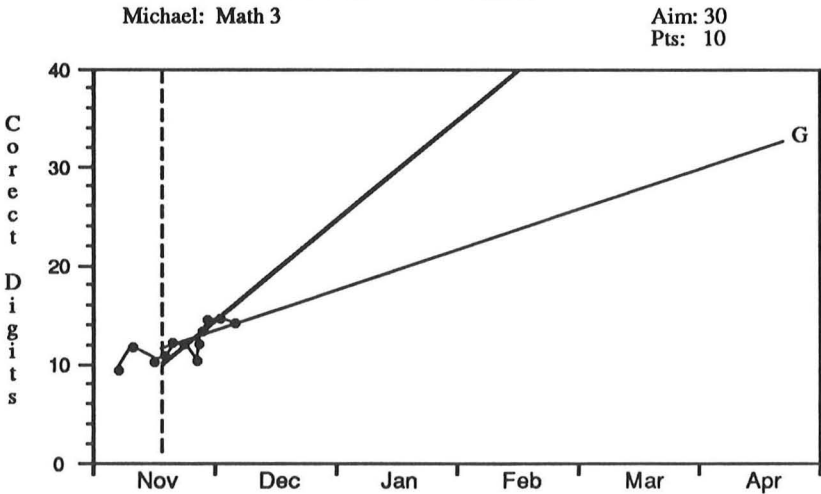
Nice work! Raise your goal.

Figure 5. Example of CMB graphs, where actual performance exceeds the progress
           anticipated by the teacher during goal setting. In the top panel, a static goal
           decision is shown, which suggests that the teacher continue to collect data.
           In the bottom panel, a dynamic goal decision is shown, which suggests that
           the teacher raise the goal.

student's rate of progress was acceptable with respect to goal attainment, and that the corresponding instructional program looked effective. The message indicated that the teacher should maintain the current instructional program and continue data collection. However, the teacher also was required to raise the goal. Figure 5 (bottom panel) shows a sample graph, illustrating satisfactory progress, with the message that corresponded to the dynamic goal CBM condition. By raising the goal, the teacher accomplished two things. First, she always adjusted the goal to correspond to the student's actual rate of progress or better; the goal was not allowed to reflect a progress rate lower than that which the student could achieve. Second, and perhaps more important, by adjusting the goal upward, the teacher was simultaneously establishing a more ambitious criterion for subsequent decisions concerning the adequacy of student progress and the instructional program. With a raise in the goal, the likelihood increased that the teacher would receive a recommendation for a teaching change in subsequent evaluations.

Two types of outcomes associated with this study are especially interesting. One type of outcome concerns teachers' use of goals; the other, student achievement. With respect to use of goals, teachers in the dynamic goal CBM group made more goal increases than teachers in the static goal CBM group. Given the dimensions of the different CBM conditions, this finding is not surprising. What is more interesting is the magnitude of effect. Within the dynamic goal group, teachers made an average of .60 goal increases; that is, they increased goals for more than one out of every two pupils. In the static goal group, only one teacher, for one of her pupils, spontaneously increased a goal in response to the student's data.

This finding is important for several reasons. First, it suggests that, despite the potential importance of ambitious goals, special educators' typical goal-setting standards may underestimate many students' potential. The study procedures allowed teachers to establish their initial goals freely, in line with the progress rates they deemed ambitious but realistic. However, with these initial goals, teachers in the dynamic goal group were required to increase goals for more than one out of every two pupils. This goal-increasing behavior was prompted by students exceeding the rates of progress teachers had anticipated. This goal-increasing rate, in response to students exceeding teachers' initial expectations, has been corroborated in additional studies we have conducted, in other academic areas. During the 1987–1988 school year, we used the dynamic goal condition in reading, spelling, and math. In these three academic areas, respectively, teachers were required to

increase goals for 4 out of every 10 pupils, 6.5 out of every 10 pupils, and 4 out of every 10 pupils. It appears that teachers may systematically underestimate handicapped students' potential to grow.

In addition to demonstrating that teachers' goals may underestimate potential progress rates, these findings indicate that without systematic prompting to raise goals, practitioners cannot be expected to do so. For example, among the 20 students participating in the static goal group, there was only one instance of a teacher raising a goal. Therefore, similar to research that indicates the importance of decision rules to prompt teachers to make instructional changes, it appears that decision rules prompting teachers to raise goals may be necessary.

The second major outcome of interest in the Fuchs, Fuchs, and Hamlett (1989a) study concerns student achievement. Concurrent with teachers' goal-raising behavior was differential student achievement. Students in the dynamic goal CBM group achieved better than the controls during posttesting on a standardized computation achievement test (with pretest performance controlled statistically). However, the achievement of the static goal CBM group did not exceed that of the controls. The effect size associated with the dynamic goal CBM procedures was .52, or approximately one-half standard deviation. This indicates that, in terms of the standard normal curve and an achievement test scale with a population mean of 100 and standard deviation of 15, one might expect the use of CBM with dynamic goals to increase the typical achievement outcome score from 100 to approximately 107.5. This finding supports previous research in psychology indicating that adults in work settings perform better with difficult goals. Additionally, findings corroborate a post-hoc special education analysis (Fuchs et al., 1985) where teachers who employed more difficult CBM goals effected better student achievement.

The Fuchs et at. (1989a) study, therefore, contributes to the CBM literature by providing an example of a workable methodology the special education community might employ for empirically deriving ambitious, but realistic, goals. A persistent problem for special education has been that during the IEP development process, before the efficacy of special education intervention has been established for a particular student, it is difficult, if not impossible, to anticipate the scope of attainable, but ambitious, goals. The Fuchs et al. study provides a process by which goals can be developed dynamically, so that progress toward mastery is monitored closely and goals are adjusted upward whenever possible. Given the finding that such goal adjustment, specifically, and goal ambitiousness, generally, may enhance student

achievement, the special education community might consider adoption of CBM systems that incorporate dynamic goal-setting procedures.

Using CBM to Judge the Adequacy of Student Progress and to Adjust Instructional Programs

Using CBM to monitor the appropriateness of instructional goals and to adjust goals upward whenever possible represents one means by which CBM can be used to assist teachers in their instructional program development. A second key way in which the CBM database can be used to enhance instructional programs is to provide the essential information with which teachers can determine (a) the adequacy of student progress, (b) the effectiveness of the current instructional program, and (c) the relative efficacy of alternative programmatic components.

Each CBM score is a performance indicator, representing the student's overall proficiency in curriculum on which measurement is conducted. Increasing scores indicate enhanced proficiency; decelerating or flat scores signify a lack of growth. As discussed previously in this chapter, when a teacher sets a goal and thereby establishes a moving goal line for a particular student, he/she simultaneously sets a minimally acceptable rate of improvement for the student, as indexed by the performance indicators. Consequently, when a student's actual rate of growth (see solid diagonal line in Figure 4) is flatter than the student's anticipated rate of growth (see broken diagonal goal line in Figure 4), a student's growth rate and the student's instructional program are judged inadequate. At this point, a recommendation is provided to make a teaching change, in order to stimulate better growth.

A series of studies indicates the importance of this "instrumental" use of the CBM database to assist teachers in judging the adequacy of student progress in order to develop enhanced instructional programs as necessary. For example, in a meta-analysis of systematic formative evaluation studies, Fuchs and Fuchs (1986) found that the use of decision rules to stimulate teachers' use of monitoring databases for programmatic development resulted in better student achievement. Fuchs et al. (1988) found a relation between student achievement and teachers' compliance with decision rules requiring teaching changes when student rates of progress were inadequate.

Additionally, in a post-hoc analysis of teachers' use of CBM in reading, Fuchs, Fuchs, and Hamlett (1989b) identified differential patterns of student achievement associated with teachers' instrumental use of CBM databases in order to formatively develop better instructional

programs. During the 1986--1987 school year, 29 teachers were assigned randomly to two treatment groups: a control group and a group that used CBM to monitor their students' reading growth. In the control group, 17 mildly handicapped students participated; in the CBM group, subjects were 36 students with mildly handicapping conditions.

In the control group, teachers used conventional special education practice to monitor student growth. As indicated on a posttreatment questionnaire, this conventional practice included unsystematic observation of student performance during lessons and grading of worksheets and other assignments.

The CBM teachers monitored student progress using CBM. Specifically, they identified curriculum levels in which student progress would be monitored and set a performance criterion for acceptable performance at the end of the 15-week study. Twice each week, teachers measured student performance with CBM. One half of the CBM teachers used a standard recall measure to monitor student growth; the other half, a standard cloze task. Additionally, within each type of measurement group, one half of the teachers measured student performance by hand and entered student scores into a data-management program; for the other half, student measurements were collected and scored automatically by computers and scores were saved directly for the data-management disk. Preliminary analyses indicated no effects associated with the type of measure condition or the type of administration factor.

Each week, teachers employed data-management software (Fuchs et al., 1987) that automatically stored and graphed the student scores, applied a set of CBM decision rules to the graphed database, and communicated decisions to teachers based on the CBM decision rules. As in the Fuchs et al. (1989a) study, the decision rules were as follows: If the student's actual rate of improvement was less steep than the goal line, the decision was to initiate an instructional change; if the student's actual rate of progress was steeper than the goal line, the decision was to increase the goal.

Following the completion of the 15-week study, the graph of each CBM student was inspected to create two CBM implementation groups: the measurement-alone group and the measurement-with-evaluation group. For the purpose of creating these two CBM subgroups, measurement was defined as administering, scoring, and graphing the curriculum-based measures on a routine basis. Evaluation was defined as the teacher introducing at least one instructional modification in response to the database and maintaining that modification for at least 2.5 weeks. Maintenance of the modification was included as a criterion

to insure that an instituted modification was in effect long enough to influence student performance.

Students were placed in the measurement-alone CBM group when their graphs showed that, although CBM measurement had occurred, the CBM database had not been used to evaluate the effectiveness of instruction and no instructional changes had been introduced in order to enhance student learning. For these students, only one viable, unchanging instructional phase had been implemented over the 15-week study. In this measurement-alone group, there were 15 students, involving nine teachers.

The remaining 21 students were placed in the measurement-with-evaluation CBM group. These students' graphs showed both that CBM data had been collected and that teachers had used the databases to evaluate and enhance instructional effectiveness. Among these students, six had three viable, different instructional phases, each implemented for at least 2.5 weeks, and 15 had two viable, different instructional phases, each implemented for at least 2.5 weeks.

Figure 6 shows two sample graphs. In the top panel, the vertical lines on the graph indicate that the teacher responded the CBM database to determine the adequacy of student growth and to develop better instructional programs; this graph would have been placed in the measurement-with-evaluation group. The bottom panel shows similar data, but the graphs lack vertical lines (i.e., no instructional changes were instituted in response to the database). Yet, as can be seen, the data pattern indicates that the teacher should have (but failed to) responded to the data instrumentally to introduce instructional changes. This graph would have been placed in the measurement-only group.

Two types of measures were used to compare the achievement of the two CBM implementation and the control groups. The first measure was a well-accepted, broadly used outcome, the Stanford Achievement Test's Reading Comprehension subtest, which was administered on a posttreatment basis and for which scores were statistically controlled using a recall measure that had been administered prior to the study. The second measure was the slope of the actual CBM database, or the rate of weekly increase in the CBM scores collected by the teachers or computers.

Results corroborated the importance of the evaluation component of CBM for effective instructional programming. Although teachers in both implementation groups set up their measurement systems and actually measured student performance using CBM comparably well, as indexed on the fidelity of treatment measure, important differences

Michael: Reading   Aim: 20
                                                                      Pts: 5

Wait: Not enough data for analysis.

Michael: Reading   Aim: 20
                                                                      Pts: 5

Uh-oh! Make a teaching change.
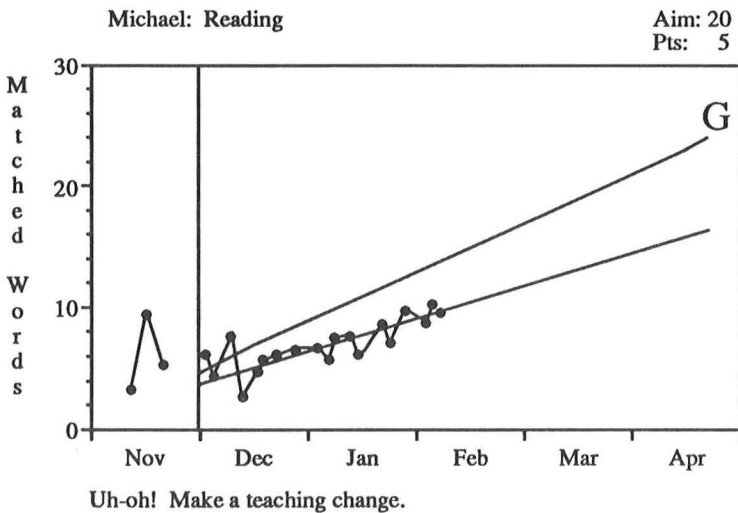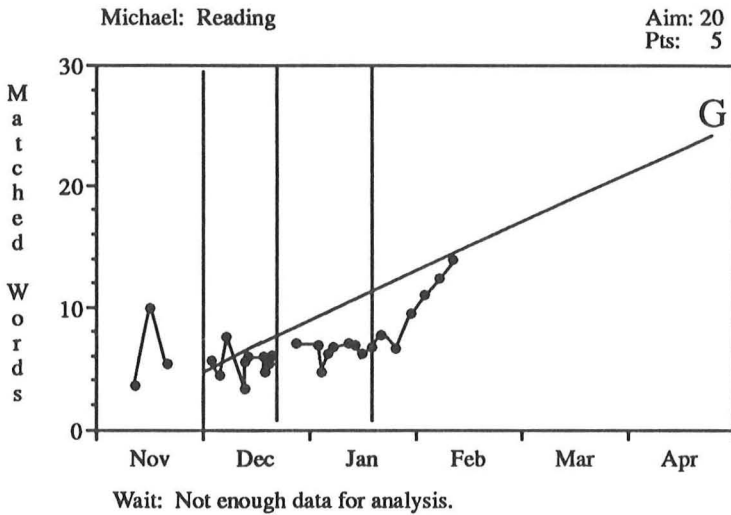
Figure 6.   Example of CBM graphs. Top panel indicates that the teacher has used the
databases to formulate instructional decision, as indicated by the vertical
intervention lines. The bottom panel shows similar data; however, the
teacher has not used the database to determine when to introduce teaching
changes in order to effect greater student growth.

were associated with the CBM implementation groups.

In terms of the global, widely accepted reading comprehension measure (the Stanford Achievement Test), findings indicated that, when teachers implemented both the measurement and evaluation components of CBM, their students achieved better (in terms of regressed adjusted scores) than the control group students. However, when teachers implemented only the measurement component of CBM, without using the database to determine when instructional improvements were warranted, student achievement did not reliably exceed that of the control group. Further, the effect size for the measurement-with-evaluation CBM group was twice as large as that of the measurement-only group.

Additionally, although the difference between the measurement-only and the measurement-with-evaluation CBM groups was not reliably different on the global Stanford Achievement Test, differences on the more direct CBM index indicated that the measurement-with-evaluation group's achievement did exceed that of the measurement-only group. The effect size was .86.

Consequently, findings support the importance of the evaluation component of CBM. With the CBM evaluation component, teachers can determine when student rates of progress are less than adequate and when program changes are warranted. When teachers not only collect CBM data, but also use CBM indicators of student growth to evaluate the effectiveness of instructional programs and to experiment with alternative instructional components, student achievement appears to be enhanced.

Using CBM to Determine the Nature of Effective Instructional Modifications

As discussed, the first strategy for using CBM databases in order to enhance teachers' instructional planning involves relying on the graphed performance indicators to monitor the appropriateness of the student's goal and to adjust the goal upward whenever necessary to ensure appropriately ambitious goals. The second strategy also involves use of the graphed performance indicators; this time, the teacher uses the graphed database to determine the adequacy of student progress and to decide when programmatic improvements appear warranted.

For both these purposes, the CBM performance indicators are employed. The performance indicators, which provide an overall index of the student's proficiency on the year-long curriculum, are well suited for summarizing the overall rate of student improvement and for

making related evaluation decisions, such as judging the appropriateness of the goal and the adequacy of student progress.

Nevertheless, the CBM performance indicators displayed on the student's graph provide relatively little direction for determining the nature of potentially effective program changes. By inspecting the performance indicators to determine the overall rate of growth in the curriculum, the teacher may be able to formulate certain potentially effective instructional changes. For example, with a flat or decelerating slope, hypotheses about (a) the lack of student retention of skills and/ or (b) motivation problems can be generated, and related programmatic changes can be considered. However, since the performance indicators do not identify which skills the student currently is performing well and which curricular components the student is not performing proficiently, the practitioner cannot use the performance indicators to formulate decisions about what dimensions of the curriculum might represent an appropriate instructional focus over the next several weeks.

Although the *graphed performance indicators* cannot be used to derive a skills profile on the target curriculum for a given student, the CBM database *does* contain the information required to put together such a skills profile. Since, during CBM testing the student is required to perform skills representing the entire year-long curriculum, student performance on all the curricular content for the year is available for each skill, on any one probe (in math, for example) or across probes (in spelling, for example). Information can be aggregated across probes to formulate a skills analysis of the student's performance.

During the 1987--1988 academic year, Fuchs and associates undertook a series of studies investigating teachers' use of the CBM skills analysis. One study was conducted in math, one in reading, and one two-part study in spelling. The studies all contrasted different types of CBM analyses teachers received to facilitate their instructional decision making. In each study, there was a control group that did not use CBM; a CBM group that relied only on the graphed database, with the related analyses to judge the appropriateness of the goal and the adequacy of student progress; and a CBM group that used both the graphed analyses as well as skills analyses that provided a skills profile to assist the teacher in determining directions for teaching changes. What follows is a detailed description of the methodology, skills analysis procedures, and results for the series of spelling studies, along with a brief description of findings in reading and math.

Spelling Study 1. Within the first spelling study, 30 special education teachers were assigned randomly to three groups: control, CBM with graphed analysis, and CBM with graphed analysis plus skills analysis. Each teacher selected two mildly handicapped pupils with spelling IEP goals to participate in the 15-week study. Analyses indicated that teachers and students in the three treatment groups were comparable on demographic variables, including (a) teachers' age, years teaching, years in current position, previous years experience in CBM research projects, highest educational degree, and personal and general teaching efficacy; and (b) students' age, grade, spelling grade level, years in special education, keyboarding skills, handicapping condition, sex, and IQ.

The control teachers in this study implemented their normal procedures for monitoring student progress in spelling. This did not include any use of CBM. As reported by the teachers in posttreatment questionnaires, the control monitoring information primarily consisted of inspection of scores on weekly quizzes assessing student proficiency on weekly spelling lists.

Within the CBM groups, teachers used CBM to monitor their two pupils' progress toward spelling goals. To establish goals, teachers (a) identified the curriculum and the level within the curriculum on which they hoped the student would be proficient by the end of the year, and (b) selected a performance criterion for acceptable performance at the conclusion of the study on April 14.

To monitor student progress toward the performance criterion of the target level of the curriculum, teachers used CBM methodology (Mirkin et al., 1984), in conjunction with computer applications (see Fuchs, Fuchs, & Hamlett, 1988). Each test was created, administered, and scored in the following way. The computer randomly sampled 20 words from the pool of words representing the target level of the spelling curriculum, and printed a hard copy of the 20-word list. A cross-age or peer tutor, aide, or teacher dictated the words from this list, and the student typed the words into the computer, with a maximum of 15 seconds before the computer automatically advanced the student to the next word. If the student finished the word before the 15-second limit, he/she pressed return to advance the computer to the next word. At the end of 20 words or 3 minutes, whichever occurred first, the computer terminated administration of the test and scored the number of correct letter sequences and words. The computer presented these scores to the student, along with a graph showing the numbers of correct letter sequences over time.

Spelling performance was measured in this way at least two times per week. Once each week, teachers used data-management software to inspect the CBM database. This software displayed a graph of the student's number of correct letter sequences over time. This graph also showed (a) broken vertical lines to represent goal changes, (b) solid vertical lines to indicate intervention changes, (c) a "G" to signify the performance criterion expected on April 14, (d) a broken diagonal line to show the goal line, and (e) a solid diagonal line to represent the student's actual rate of progress.

The computer applied the following set of decision rules to the graphs. If the student's actual rate of progress was steeper than the goal line, a decision appeared below the graph saying, "Nice work! Raise your goal." If the student's actual rate of progress was flatter than the goal line, a decision read, "Uh-oh! Make a teaching change." If the student's recent scores were higher than a predetermined ceiling level, a decision read, "Move to the next curriculum level." Finally, if there were fewer than eight new scores since the last vertical line, the decision read, "Insufficient data. Keep collecting data." The computer used an interactive structure to communicate these decisions (see Fuchs, Fuchs, & Hamlett, 1988), where teachers had to inspect the database independently and enter their own decisions. The computer provided corrective feedback to the teachers' responses and provided explanations for correct decisions (see Fuchs, Fuchs, & Hamlett, 1988). CBM teachers in the graphed analysis and in the graphed plus skills analysis received this graphed feedback.

CBM teachers in the graphed plus skills analysis group, however, received additional information. Using the most recent 50 words the student had spelled, the computer provided the following skills analysis. The computer indicated the number of correctly spelled words, the number of Near Misses (incorrect words with at least 50% correct letter sequences), and the number of Far Misses (incorrect words with fewer than 50% correct letter sequences). The computer also identified, for every word in the Near Misses category, the error categories the student had committed, and then showed the teacher (a) for each possible error type, the number of corrects and opportunities, as well as the percentage correct, and (b) three key error categories the student had made most frequently, along with up to four examples of each frequent error category. Finally, the computer presented the teacher with complete lists of the Corrects, Near Misses, and Far Misses. Figures 7 and 8 show a sample 2-page printout of the information contained in the spelling skills analysis.

Spelling Profile

Name:	Domain: Spelling D	Date: 4/15/89	Page 1

Corrects (100 LS Correct):	14 words(s)
Near Misses (60-99% LS Correct):	19 words(s)
Moderate Misses (20-59% LS Correct):	16 words(s)
Far Misses (0-19% LS Correct):	1 words(s)

Type	Correct	Possible	PCT
FSLZ	0	0	--
Final E	1	5	20
Blend	7	10	70
Double	3	4	75
Dual Con	12	24	50
Vowel+R	9	14	64
Vowel+N	6	8	75
Suffix	4	6	67
Digraph	7	10	70
Vow Team	4	12	33
C/S	0	1	0
-Le Word	4	7	57
Final Vow	3	7	43
Ild Word	0	0	--
Dge Word	0	1	0
Ch/tch	2	2	100
C/ck	0	2	0
Shun Word	0	1	0
Combo	1	1	100
Ign/igh	0	0	--
V+L+Con	0	0	--
Sure Word	0	0	--
Ance Word	0	0	--
Irregular	1	1	100
Apos'Phe	0	0	--
Sing Vow	21	30	70
Sing Cons	47	49	96

Key Errors

Vow Team	Dual Con	Final E
Instead-Insted	Learner-Leaner	Alone-Alon
Moisten-Mosten	Sample-Samble	Knife-Knif
Quieter-Quiter	Chart-Chard	Rare-Rar
Trouble-Trubble	Mumble-Mobble	Cube-Cub
Rail-Real	Tractor-Trater	
Certain-Chanten	Apart-Apeot	

Figure 7. Page 1 of the computerized CBM spelling skills analysis.

```
------------- Moderate Misses (20-59% LS Correct) -----------------
```

57 Tickle-Teakle	C/CK	Sing Vow		
57 French-Fanch	Vowel + N	Blend		
57 Mumble-Mobble	Dual Con	Sing Vow		
50 Unlucky-Unluke	Final Vow	C/CK		
50 Tractor-Trater	Vowel + R	Dual Con		
50 Apart-Apeot	Vowel + R	Dual Con		
44 Calendar-Cander	Vowel + R	Vowel + N	Sing Cons	
42 Mumble-Mommbe	-Le Word	Sing Vow		
40 Rail-Real	Vow Team			
37 Station-Stanch	Shun Word			
28 Sample-Scembe	-Le Word	Dual Con	Sing Vow	
25 Certain-Chanten	Vow Team	Vowel +R	C/S	
25 Squeeze-Sceasc	Vow Team	Digraph	Digraph	
20 Limb-Lem	Dual Con	Sing Vow		
20 Treatment-Tempemt	Suffix	Vow Team	Blend	
20 Limb-Leam		Dual Con	Sing Vow	

```
------------------------- Far Misses (0-19% LS Correct) ----------------
```

14 Giggle-Gelly	-Le Word	Double	Sing Vow

Figure 8. Page 2 of the computerized CBM spelling analysis.

Several types of outcome measures were collected. First, fidelity of treatment was indexed. Second, teachers' program development was measured in several ways. Finally, student achievement was assessed using a standardized spelling achievement test, which required students to write Grades 1-6 words that appear with high frequency across curricula. Results indicated the following.

With respect to fidelity of treatment, teachers in the two CBM groups structured their measurement procedures and actually measured student performance in a highly accurate and comparable manner. However, teachers in the graphed-plus-skills-analysis group received relatively high fidelity of treatment scores for the Evaluation component of the fidelity of treatment scale; their Instructional Plan Sheets, on which they recorded their teaching changes, were completed in a more acceptable fashion, compared to the graphed-analysis-only teachers.

In a related way, for program development, teachers in the two CBM groups scored comparably on most variables, including number of goal increases, level of goal ambitiousness, and number of teaching changes. However, teachers in the graphed-plus-skills-analysis group received higher scores than teachers in the graphed-analysis-only group on the number of skills they targeted for instruction and listed on their Instructional Plan Sheets.

In terms of achievement, teachers in the graphed-plus-skills-analysis group effected greater growth compared to (a) teachers in the graphed-analysis-only group and (b) teachers in the control group. The average gains from pre-to posttesting for the graphed-plus-skills-analysis group, the graphed-analysis-only group, and the control group, respectively, were approximately 37, 14, and 12.

Consequently, it appeared that the skills analysis information contributed critical information in order to promote effective instructional planning. With the addition of the skills analysis to the graphed feedback, teachers were able to write more acceptable instructional programs; they cited more skills to target during instruction; and they effected superior student achievement. Results of this study strongly support the usefulness of skills analysis within CBM to support teachers' effective instructional decision making.

Nevertheless, an important shortcoming of this study, with respect to generalization to typical CBM procedures, is that the graphed-analysis-only procedures used in this study involved computerized data collection. This meant that teachers did not routinely inspect students' spelling performance. Yet, with typical CBM, which does not rely on automatic data collection, teachers frequently score and thereby inspect student spelling samples. With computerized data collection,

however, teachers do not routinely score student tests. Rather, they typically see *only* the graphed analysis. Because of this limitation associated with the computerized data collection used in this study, a second, related investigation was undertaken. (For a complete description of this study, see Fuchs, Fuchs, Hamlett, & Allinder, in press.)

Spelling Study 2. In this second study, the 30 same teachers were assigned randomly to three treatment groups: control, CBM with graphed-plus-skills analysis, and CBM with graphed analysis plus Near Misses inspection. Study procedures were identical to those employed in Study 1, with the following deviation. This time, CBM teachers who did not receive the skills analysis did have the opportunity to inspect student spellings. This was accomplished in the following way. After viewing graphs and receiving the graphed analysis, teachers in the graphed analysis plus Near Misses inspection group saw the list of Near Misses. The Near Misses list contained incorrectly spelled words from the pool of the most recent 50 words the student had spelled on his/her tests. These Near Misses had to be at least 50% correctly spelled, in terms of letter sequences. They were presented to the teachers from most correct (99% letter sequences correct) to least correct (50% letter sequences correct), with the correct and incorrect spellings next to each other. (See page 2 Near Misses of Figure 8; however, only the correct and incorrect spelling were provided in this Near Misses treatment.)

This Near Misses condition was incorporated into Study 2 in order to provide teachers, who did not receive formal skills analysis, an opportunity to view a structured presentation of the student's spelling errors. This structuring of the student's Near Misses provided richer information than the graphed analysis only condition of Study 1 and therefore better approximated typical CBM procedures where teachers score student tests by hand. Nevertheless, the Near Misses condition provides a more systematic and structured presentation of information than is inherent within the simple hand scoring teachers complete with noncomputerized CBM. Consequently, the Near Misses condition must be viewed as a form of CBM that presents teachers with information somewhat less organized than skills analysis but more systematic than provided by simple hand scoring.

Results of this second study indicated the following. CBM teacher performance was comparable on fidelity of treatment and program development indices. However, teachers did effect differential achievement among their students. Progress for the students within the graphed-plus-skills-analysis groups was reliably better than that of

controls (an average gain of approximately 33 versus approximately 12). However, the difference in achievement between the Near Misses group and the control-only group approached statistical significance (p = .07), with mean gains of approximately 24 versus 12. The difference in growth between the skills analysis and the Near Misses group was not reliably different. (For a complete description of this study, see Fuchs, Allinder, Hamlett, & Fuchs, in press.)

This series of studies suggests the following. First, skills analysis does seem to provide teachers with structured information that supplements the graphed CBM database in such a way that facilitates teachers' effective instructional decision making. Second, as additional sources of structured feedback are provided to teachers (graphed analysis vs. Near Misses lists vs. skills analysis), teachers' instructional decision making and student achievement appears to be enhanced.

Reading and math studies. During the 1987-1988 academic year, similar studies were conducted in the areas of reading and math. In these additional academic areas, CBM teachers either received graphed feedback only or graphed feedback with skills analysis. In both additional academic areas, results were similar to those found in spelling. That is, with the additional information supplied by the skills analysis, teachers were able to structure better instructional programs and they effected superior student achievement. Consequently, the finding that teachers can use additional sources of feedback about student performance, including skills analysis, to enhance instructional decision making appears to be robust. (For descriptions of the reading and math studies, respectively, see Fuchs, Fuchs, & Hamlett, 1989 and Fuchs, Fuchs, Hamlett, & Stecker, 1990.)

Concluding Remarks: Getting Teachers to Use CBM

This review of research highlights three ways in which teachers may use CBM databases to assist in their instructional decision making: (a) to monitor the appropriateness of their goals and to adjust goals as necessary, (b) to judge the adequacy of student progress and to create instructional modifications when needed, and (c) to rely on skills analysis to derive additional information from the CBM database for formulating potentially effective instructional improvements.

As noted, studies have documented that CBM can be used to effect statistically significant and practically important differences in student achievement outcomes across academic areas. Yet, as noted by Wesson, King, and Deno (1984) and others (e.g., Walton, 1986), teachers are reluctant to employ CBM and other forms of ongoing student

performance monitoring, because these measurement systems are time consuming and frequently technically demanding (see Wesson, Fuchs, Tindal, Mirkin, & Deno, 1986).

A pressing question, then, is: How can we facilitate teachers' implementation of ongoing assessment systems and induce teachers to use these systems effectively? Our CBM intervention research suggests the following. First, computers can be used to reduce teacher time necessary to implement CBM. With computerized automatic data collection in reading, spelling, and math (Fuchs et al., 1990), the teacher is freed from the time-consuming tasks of developing measures, administering and scoring tests, and analyzing student performance profiles. Rather, once students have been taught to use the CBM software, teachers need only to view assessment profiles (i.e., graphs and skills profiles that are produced automatically by computers). Evidence indicates that with these automatic data collection and analysis programs, teacher time devoted to measurement can be virtually eliminated and teacher satisfaction with CBM improves (Fuchs, Hamlett, Fuchs, Stecker, & Ferguson, 1988).

Despite this improved feasibility, it appears that teachers may still require some inducement to incorporate the information presented in CBM assessments into their instructional decision making. Research (e.g., Fuchs, Fuchs, Hamlett, & Ferguson, 1989; Tindal, Fuchs, Mirkin, Christenson, & Deno, 1981) indicates that teachers may experience difficulty in formulating effective strategies for revising their instruction when student performance data indicate that student rates of progress are inadequate. Additionally, given the increasing numbers of students on many special education roles and the complexity and diversity of class compositions in regular and special education settings, the individual nature of the CBM assessment profiles and instructional implications may be problematic for teachers. That is, teachers may recognize not only *when* they need to revise different students' programs, but also *how* they might improve student programs. Yet, the numbers and types of students and the many *different* instructional adaptations indicated by the CBM data may preclude or reduce the likelihood of teachers' responsive use of a CBM database.

In our CBM research we have tried to address these two problems (i.e., teachers' need for assistance in formulating potentially effective revisions to their students' instructional programs and the logistical difficulties in revising different students' programs in different ways at different times), in several ways. First, in terms of support to teachers in order to assist them in formulating potentially effective instructional

revisions, consultants (i.e., our project staff) visit teachers once every 1-2 weeks, review with them the CBM student profiles, and assist them in identifying instructional revisions, including the provision of instructional packets to assist teachers in specifying and implementing instructional modifications.

Second, as a alternative to frequent consultant visits, we have developed and researched computerized expert systems that provide systematic consultation in reading, spelling, and math. Our initial research (e.g., Fuchs, Fuchs, Hamlett, & Ferguson, in press) using these computerized recommendation systems indicates that they may represent an effective substitute for the relatively expensive use of consultants.

Third, with respect to the logistical problems of implementing many programmatic changes for different students at different times, we have begun to develop and research computer programs that simultaneously consider all students on an individual teacher's caseload. These programs present information and make instructional suggestions for flexible groupings of students, rather than for individuals. We hope that with these group profiles and recommendations, teachers will revise instructional groupings more frequently and implement sound instructional strategies for these flexible student groupings. Research investigating this possibility is under way.

REFERENCES

Anderson-Inman, L., Walker, H., & Purcell, J. (1984). Promoting the transfer of skills across settings: Transenvironmental programming for handicapped students in the mainstream. In W. L. Heward, T. E. Heron, D. S. Hill, & J. Trap-Porter (Eds.), *Focus on behavior analysis in education* (pp.17-37). Columbus, OH: Merrill.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*, 122-147.

Blankenship, C. (1985). Using curriculum-based assessment to make instructional decisions. *Exceptional Children, 52*, 233-238.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning.* New York: McGraw-Hill.

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.

Fuchs, L. S., Allinder, R. M., Hamlett, C. L., & Fuchs, D. (1990). An analysis of spelling curricula and teachers' skills in identifying error types. *Remedial and Special Education, 11* (1), 42-53.

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,* 449-460.

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

Fuchs, L. S., & Fuchs, D. (1990). Curriculum-based assessment. In C. Reynolds & R. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children (Vol. 1): Intelligence and achievement* (pp. 435-455). New York: Guilford Press.

Fuchs, L. S., Fuchs, D., & Deno, S. L. (1985). The importance of goal ambitiousness and goal mastery to student achievement. *Exceptional Children, 52,* 63-71.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1988). Computers and curriculum-based measurement: Effects of teacher feedback systems. *School Psychology Review, 18,* 110-123.

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

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989b). 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., (in press). Monitoring reading growth using student recalls: Effects of two teacher feedback systems. *Journal of Educational Research.*

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. (in press). The contribution of skills analysis to curriculum-based measurement in spelling. *Exceptional Children.*

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (in press). Effects of expert system advice within curriculum-based measurement using a reading maze task. *Exceptional Children.*

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

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1990). *Basic Reading, Basic Spelling, Basic Math.* [Computer Programs.] Austin: PRO-ED, Inc.

Fuchs, L. S., Hamlett, C. L., Fuchs, D., Stecker, P., & Ferguson, C. (1988). Conducting curriculum-based measurement with computerized

data collection: Effects on efficiency and teacher satisfaction. *Journal of Special Education Technology, 9* (2), 73-86.

Gagne, R. M. (1964). The implications of instructional objectives for learning. In C. M. Lindvall (Ed.), *Defining educational objectives.* Pittsburgh: University of Pittsburgh Press.

Gickling, E. E., & Thompson, V. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52,* 205-218.

Goodstein, H. A. (1982). The reliability of criterion-referenced tests and special education: Assumed versus demonstrated. *The Journal of Special Education, 16,* 37-48.

Hartley, J., & Davies, I. K. (1976). Preinstructional strategies: The role of pretests, behavioral objectives, overviews, and advance organizers. *Review of Educational Research, 46,* 239-265.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. (1978). *Measuring pupil progress toward the least restrictive environment* (Monograph No. 10). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 185 767)

Mager, R. F. (1975). *Preparing instructional objectives.* Belmont, CA: Fearon Publishers.

Marston, D., Fuchs, L. S., & Deno, S. L. (1985). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-related measures. *Diagnostique, 11,* 77-90.

McNeil, J. D. (1967). Concomitants of using behavioral objectives in the assessment of teacher effectiveness. *Journal of Experimental Education, 35,* 69-74.

Mirkin, P. K., Deno, S. L., Fuchs, L. S., Wesson, C., Tindal, G., Marston, D., & Kuehnle, K. (1984). *Procedures to develop and monitor progress toward IEP goals* (2nd ed.). Minneapolis: University of Minnesota.

Peckham, P. D., & Roe, M. D. (1977). The effects of frequent testing. *Journal of Research and Development in Education, 10* (3), 40-50.

Rosswork, S. G. (1977). Goal setting: The effects on academic tasks with varying magnitudes of incentive. *Journal of Educational Psychology, 69,* 710-715.

Salvia, J., & Hughes, C. (1990). *An introduction to the assessment of achievement.* New York: Macmillan.

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

Shinn, M. R., Rosenfield, S., & Knutson, N. (1989). Curriculum-based assessment: A comparison of models. *School Psychology Review, 18,* 299-316.

Tindal, G., Fuchs, L. S., Fuchs, D., Shinn, M. R., Deno, S. L., & Germann, G. (1985). Empirical validation of criterion-referenced tests. *Journal of Educational Research, 78* , 203-209.

Tindal, G., Fuchs, L. S., Mirkin, P. K., Christenson, S., & Deno, S. L. (1981). *The effect of measurement procedures on student achievement* (Research Report No. 61). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 218 847)

Tucker, J. (1987). Curriculum-based assessment is no fad. *The Collaborative Educator, 1*(4), 4, 10.

Walton, W. T. (1986). Educators' responses to methods of collecting, storing, and analyzing behavioral data. *Journal of Special Education Technology, 7,* 50-55.

Wesson, C., Fuchs, L. S., Tindal, G., Mirkin, P. K., & Deno, S. L. (1986). Facilitating the efficiency of on-going curriculum-based measurement. *Teacher Education and Special Education, 9,* 166-172.

Wesson, C., King, R., & Deno, S. L. (1984). Direct and frequent measurement: If it's so good for us, why don't we do it? *Learning Disability Quarterly, 7* (1), 45-48.

White, O. R. (1984). Descriptive analysis of extant research literature concerning skill generalization and the severely/profoundly handicapped. In M. Boer (Ed.), *Investigating the problem of skill generalization: Literature review* (pp. 1-19). Seattle: University of Washington, Washington Research Organization.

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

Yalow, E. S., & Popham, W. J. (1983). Content validity at the crossroads. *Educational Researcher, 12* (8), 10-14, 21.

4

Academic Skill Assessment: An Evaluation of the Role and Function of Curriculum-Based Measurements

Francis E. Lentz and Jack J. Kramer

University of Cincinnati and University of Nebraska-Lincoln

In the most meaningful use of the term assessment, important decisions are made daily by teachers based on their assessment of information obtained from student responses to curriculum-related materials. These assessment decisions may include deciding on extra work or deciding to refer a child for learning or behavior problems. The term curriculum-based assessment (CBA) has been used to encompass a wide range of procedures ranging from these daily informal analyses by teachers, to highly structured measurement systems

Authors' Notes. This chapter and the presentation by the first author at the Buros-Nebraska symposium were based in part on material previously published elsewhere (Lentz, 1988).

used in special education systems. Although well-constructed guides exist for some sets of curriculum-based decisions (e.g., Shinn, 1989), there is inadequate empirical research to assist our understanding of how, or how well, most of these decisions are made.

Recently, attempts have been made to formalize the use of measures of student academic performance, especially in decisions about special education eligibility for students who seriously fail to meet classroom expectations (i.e., Tindal, 1988; Shinn, 1989). At least one type of CBA developed for special education systems, called curriculum-based measurement (CBM) has been the subject of extensive evaluation research (see Tindal, 1988, for a comprehensive review) and interest on the part of special service personnel such as school psychologists (e.g., Shapiro, 1990) and special educators (e.g., Tucker, 1985). Yet, as interest has grown many questions have arisen about what we know about CBA, and we think more importantly, about *how* we know what we know!

With this paper we have set modest goals. It will be suggested that curriculum-based assessment fits best within a behavioral model of measurement and an examination of that assumption is provided. The discussion of the behavioral assessment model provides a foundation for our review of curriculum-based assessment (CBA) and the manner in which CBA has been developed and used. The approach taken herein is to some degree critical based on our analysis that many questions remain unanswered, questions about the nature of curriculum-based measures themselves and the manner in which the emerging CBA technology has been and will be applied. However, we wish to strongly emphasize our belief that CBA has already had a positive influence on educational practice, especially our understanding of how to help teachers make better decisions in order to enhance academic achievement (see, for example, Fuchs, this volume), and has served an equally important heuristic influence on the field of educational measurement.

We think CBA potentially has much more to offer in improving measurement within the assessment of school based problems. Our analysis suggests that CBA is best understood not as a monolithic assessment procedure, but as a source of data to be considered along with other sources in a comprehensive analysis of academic skills and learning environments. Because of this, CBA must be evaluated as part of, not different from, the entire evaluation process. To date this has rarely been accomplished (see Lentz, 1988, for an exception). We will argue that a choice of specific procedures (e.g., CBA, standardized intelligence or achievement tests, event sampling) to be used during an

assessment should flow from an understanding both of the general assessment model to be followed and the specific assessment questions to be answered for a particular child. In this regard we are particularly interested in the use of CBA data within intervention assistance programs for at-risk students.

There appear to be many questions about the manner in which CBA procedures should and will be implemented in classroom settings. Specifically, we are concerned about the manner in which CBA will be adopted by school psychologists and the entire educational establishment. For example, we foresee a number of problems with piecemeal adoption of structured CBA procedures by a portion of special services staff (e.g., school psychologists but not special education teachers or vice versa). We fear that in the absence of a clear assessment model or evaluation goals, CBA may be used in a manner that diverts attention from other environmental factors (e.g., instructional variables) that may contribute to academic success or failure. For example, if evaluators focus prime attention on CBA data during decision making for intervention planning, then problems may arise because of the overemphasis on student skill or fluency deficits at the expense of examining problems between students' performance and the instructional environment. Publications describing CBM use seem to continue to address placement special education issues (and subsequent IEP development or monitoring) and deemphasize intervention assistance prior to placement (e.g., Marston and Magnusson, 1988).

Public education does not have an impressive track history of adopting efficacious procedures in a timely or comprehensive manner (e.g., Bickel & Bickel, 1986; Greer, 1983) and we are concerned CBA may be ignored, or perhaps even worse, be used in a manner that perpetuates bad practice. Unfortunately, many of the problems that CBA attempts to address are not simply due to the lack of a better mousetrap. The technology for assessing behavior directly and altering response patterns of children within educational settings has been around for some time (Benes & Kramer, 1989). Even within our own profession, alternative assessment and psychological service delivery models for public schools have been suggested for many years (Gallessich, 1974; Hops, 1971), but school psychologists have not rushed to implement innovative service delivery strategies (Conoley & Gutkin, 1986). The data indicate clearly that most school psychologists know that there are more useful ways to spend their time than administering standardized tests and placing children in special class programs (e.g., Goldwasser, Meyers, Christenson, & Graden, 1983; Kramer & Peters, 1986). There are, however, many incentives for continuing the refer-test-place process.

We must guard against CBA becoming part of the systemic problems which detract from effective psychological services in schools in order to avoid attenuating the potential impact of curriculum-based (or other direct) measures of academic performance. In terms of CBA having a meaningful impact on services for the wide range of children with academic problems, the most important question may be whether CBA will have primary impact on children after they are classified, or whether CBA can become a key factor in assisting at-risk students irrespective of handicapping condition.

In summary, our objectives for this paper include: (a) examination of the behavioral assessment model and the implications of this model for educational measurement; (b) review of the development, utilization and evaluation of CBA procedures; (c) discussion of potential implementation problems with CBA; and (d) suggestions for further conceptualization, development, and implementation of CBA and other direct measures of academic behavior.

EVALUATING CBA: WHICH MEASUREMENT MODEL IS APPROPRIATE

The requirement for practitioners to evaluate and select appropriate assessment methods is clear from both ethical and professional perspectives (e.g., American Psychological Association, 1981). In this regard, a set of guidelines for appropriate test evaluation is available (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1985). However, we believe serious conceptual and practical difficulties face practitioners and researchers in making decisions in regard to selecting, evaluating, interpreting, and using specific assessment methods within an assessment process. Most "traditional" tests have been developed and evaluated using one of several psychometric models that provide frameworks for the collection of data on some quality of a specific test, rather than how useful a test is within an actual decision-making process that nearly always involves multiple information sources. For example, data may be available on the reliability of a test, but not on the reliability or stability of educational decisions made using such a test. MacMahen and Barnett (1985) have provided startling conclusions about the unreliability of decisions made using reliable tests.

Similarly, most psychometric models usually treat functional environmental influences on test performance as some sort of error. Test scores are interpreted within confidence bands derived from studies of variance in sets of test scores and standard extrapolations are

applied to individual scores. For the issue at hand, academic measurement, traditional tests are interpreted as telling us how much of some construct an individual has (reading ability, for example).

TECHNICAL EVALUATION OF CBA

Most recently, the term curriculum-based assessment (or, measurement) has been most closely associated with research conducted at the University of Minnesota (e.g., Deno, Marston, Shinn, & Tindal, 1983; Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Deno, 1982; Fuchs, Tindal, & Deno, 1984; Germann & Tindal, 1985; Shinn & Marston, 1985), and outcomes of this research have been extensively disseminated. Academic probes of 1-2 minute duration were developed from curriculum materials with the goals of efficiency, simplicity, ease of interpretation, applicability to a wide range of academic decision, and cost being central to the design of the procedures (Deno, 1985). Investigation of the use of curriculum probes has been conducted across a variety of academic skill areas including reading (e.g., Deno, 1985), spelling and writing (e.g., Germann & Tindal, 1985), and arithmetic (e.g., Blankenship, 1985). Although such brief probes were originally conceptualized as a means of progress monitoring, probes have been examined for a number of different assessment functions within the framework of special education decision making.

In his review of direct measurement of academic behavior, Lentz (1988) has examined the functions to be served through the assessment process and the contributions of CBA to each. He suggests that CBM measures have been used for: screening for program eligibility (e.g., Marston & Magnusson, 1985), placement in curriculum levels (e.g., Deno & Mirkin, 1977), and most prominently, progress monitoring (e.g., Deno, 1985). Until recently (see e.g., Fuchs, this volume), little attention has been given to using CBA systems, at least of the type developed at the University of Minnesota, in identification of specific variables as targets for intervention.

The fact that CBM investigations have produced more direct and cost efficient methods (as compared to traditional standardized testing) for eligibility decisions or monitoring educational progress cannot be denied. Indeed, the data obtained in the Minnesota investigations suggest that curriculum-based probes "are as psychometrically sound as standardized achievement tests, are much simpler to administer, and are much less expensive" (Lentz, 1988, p. 98). CBA data have been used to differentiate among exceptionalities and place children in special programs (Marston & Magnusson, 1985; Shinn & Marston, 1985).

Others have advanced methods of developing local norms for CBA data (Shinn, 1988) with the suggestion that these data can be used to assist in the identification and placement of children in special programs.

Although each of these articles address issues of interest and importance, we see much reason for concern both in the general approach suggested by this research and the specific manner in which CBA is utilized in these investigations. As discussed above, we are not comforted by the fact that CBA procedures fulfill many traditional psychometric assumptions (e.g., reliability and validity). We are just as troubled by our perception that a prime interest appears to be in the use of CBA data to assist in placement of children within special programs.

Although CBM has primarily been evaluated within a traditional psychometric model, there are several notable exceptions. Fuchs and her associates (e.g., Fuchs, 1989) have provided convincing evidence that using CBM for systematic goal setting, progress monitoring, and decision making about instructional change can enhance student achievement in reading, math, and spelling. This strand of research seems best conceptualized as research into the validity on an intervention, the intervention being making data-based decisions, and also seems most related to a behavioral assessment model.

Initial CBM research appeared to accept implicitly the premises of a traditional psychometric model, with studies of internal consistency (Fuchs, Fuchs, & Deno, 1982), test-retest reliability (Marston & Deno, 1981), and concurrent validity (Deno, 1985) predominating; however, few studies appear to have examined *decision* reliability or validity of CBM. For example, the stability of placement decisions made with CBM data across assessors, time, or even different eligibility rules have not been closely examined. Unfortunately, some data (e.g., Derr & Shapiro, 1989) have suggested that these factors may affect eligibility decisions.

There have been a number of other recommended uses of CBA that would not appear to fit within a traditional measurement paradigm. For example, Lentz and Shapiro (1986) and Shapiro (1990) have outlined the use of curriculum-based written products and CBM type probes during problem analysis for planning interventions, or in assessing environmental influences on academic problems. Likewise, Gable and Hendrickson (1990) provide guidelines for using student performance measures in specific instructional planning. However, there appears to be no empirical evaluation of these suggestions. Further, given the purpose of these suggested procedures, the traditional measurement model does not offer an appropriate framework for evaluating assessment adequacy.

The behavioral assessment model has been presented as a viable alternative to traditional trait-oriented measurement models that once dominated (Haynes & Wilson, 1979; Hersen & Bellack, 1981). During the last two decades many direct observation procedures that are conceptualized as behavioral assessment have been used in classroom research and assessment (e.g., Kazdin, 1984) and there are a number of academic assessment systems, including CBA (e.g., Deno, 1985; Haring & Eaton, 1978; White & Liberty, 1976) that to some degree correspond to the behavioral assessment model in terms of assumptions about measurement and the functions of assessment.

Traditional approaches to measurement have often used behavior as signs or signals of some underlying condition that the individual *has*, whereas behavioral assessment is more interested in the individual's actual behavior, that is, what the individual *does* (Hartmann, Roper, & Bradford, 1979; Haynes & Wilson, 1979). This reluctance to infer beyond the behavior itself or to consider behavior as a sign of some abstract construct of diffuse state is a defining characteristic of the behavioral assessment model. In addition, behavior is considered to be to some degree situationally specific and considerations of reliability and validity of assessment procedures must be made relative to actual behavior in natural settings (e.g., Cone, 1981; Hartmann et al., 1979).

The behavioral assessment model has led to the development of many measurement procedures that have found extensive application in education and psychology. Specific applications of behavioral assessment have included selection of clients, identification of target behaviors, determination of controlling variables, selection of treatment procedures, and monitoring and evaluation of treatment efficacy (Nelson & Hayes, 1981). In order to accomplish the tasks described above, behavioral assessment emphasizes direct, repeated measurement of behavior and controlling variables in the environments in which the behavior of interest occurs. Of course, it is true that the ideal of direct and repeated measurement in the environments of interest may not always be possible; however, this assessment model offers the potential for direct linkage between assessment and intervention.

EVALUATING BEHAVIORAL ASSESSMENT

Within the behavioral assessment model, measurement data have been conceptualized along several dimensions. First, data can be analogue or natural. In the former, data on actual behavior are collected, but in settings that are not naturally where the behavior occurs, for example, role-play tests. Natural data are collected within the actual

settings of interest. A second dimension is whether measures are direct or indirect. In both, behavior is of prime interest, but in the former, data are collected concurrently with the occurrence of target behavior (for example, direct observation of behavior), but in the latter, data are collected retrospectively (for example, behavior checklists).

In behavioral assessment, the accuracy of measurement (direct relationship to criteria characteristics of ongoing behavior) and relationship of data to functional controlling variables and critical behaviors in natural settings are prime criteria for evaluating measurement utility. Because a prime purpose of a behavioral assessment is to measure environmental (and other) variables that maintain current target behaviors (or inhibit acquisition of more appropriate behaviors), assessment procedures must be evaluated in terms of how well they accomplish this purpose. Only if assessment data provide such information, can intervention plans be directly linked to assessment information.

Some CBA data are direct and natural, such as work samples, curriculum embedded tests, and measures of oral responding during class activities. Other types of CBA data, for example, that included under the rubric of curriculum-based measurement (e.g., Tindal, 1988) are direct and analogue in nature; behavior is measured directly but under contrived conditions (not as part of "naturally occurring" academic behavior in the classroom). The developers of CBM seem to have conceptualized CBM probe data as a "sign" or construct of academic skills or achievement, similar to traditional achievement tests in this aspect, and to have evaluated it primarily in this traditional regard (e.g., Tindal, 1988). If CBA data are used in academic assessment oriented towards intervention planning, then evaluation of their adequacy would seem best derived from a behavioral assessment model and related assumptions. Even as used in progress monitoring (repeated measures of direct analogue measures), CBM would seem more related to the purposes of measurement within a behavioral measurement model.

CBA has not been clearly and consistently related to ecologically valid criteria (Martens & Witt, 1988). For example, are positive data series obtained through repeated CBA reading probes consistently related to improvements in children's oral reading in instructional reading groups? Do teacher's perceptions of change in the way children meet classroom expectations correspond to CBA data? When there is a lack of correspondence between CBA data and teacher perception of change (or actual classroom behavior), what then? How consistent are CBM measures gathered across different raters and different settings?

In the next section, we examine further the limitations of more traditional approaches to assessment and consider more completely the advantages of conceptualizing academic assessment and CBA within a behavioral assessment model.

ACADEMIC ASSESSMENT WITHIN A BEHAVIORAL ASSESSMENT MODEL

Trait-oriented approaches to educational measurement have not proven to be very productive. Although schools continue to spend a great deal of time assessing constructs such as intelligence and mental processes (e.g., auditory memory, simultaneous and sequential processing), the treatment utility of such approaches remains elusive (e.g., Arter & Jenkins, 1979; Kramer, Henning-Stout, Ullman, & Schellenberg, 1987; Witt & Gresham, 1985). Inferences about global tendencies (e.g., attention, impulsivity) that have often been made based on subject behavior during testing have not been shown to be any more useful than our attempts to measure intelligence or cognitive processes.

Trait-oriented procedures, relying on norm references for quantitative measurement, have been criticized in ways that are related to the differences between traditional and behavioral assessment models. Norm-referenced approaches:

> do not offer absolute measures of academic behavior; rather, the meaning of derived measures comes from a student's relative standing in a norm group. They are also difficult to use in a frequent, repeated fashion and are thus not useful for progress monitoring. The lack of direct relationship between achievement tests and what is actually taught to children has also been highly criticized . . . . (Lentz, 1988, p. 83)

As will be seen, CBM has depended on being norm referenced for a variety of purposes, including screening, placement, and goal setting (Tindal, 1988). However, because of the nature of this type of CBM measure, it appears much more sensitive to interventions, and more useful in repeated progress monitoring than standardized achievement tests. Other approaches to academic assessment also approximate the requirements of a behavioral assessment model. For example, the content of criterion-referenced tests closely resembles academic behavior required in classrooms. Although performance on a criterion-referenced test is not a direct measure of classroom responding, responses on these

tests could be considered analogue measures. A serious concern with criterion-referenced tests is that of variable quality, which further limits the extent to which these instruments approximate classroom behavior (Tindal, Fuchs, Fuchs, Shinn, Deno, & Germann, 1985).

Curriculum-related academic assessment and intervention systems have been specifically and purposefully developed to overcome many of the problems identified with norm-referenced achievement tests. For example, data-based instruction (Haring & Eaton, 1978), precision teaching (White & Liberty, 1976), and curriculum-based measurement (e.g., Deno, 1985) all assess academic skills and employ direct observation and measurement procedures. These procedures focus on academic skills, target the goals of classroom instruction, and often use materials taken directly from the classroom curriculum. They differ from criterion-referenced tests in that they involve brief, timed, and frequently administered probes of precisely defined academic behavior. As discussed above, although the measurement stimuli used in these systems are taken directly from classroom curriculum, the conditions under which stimuli are administered may not mirror natural classroom conditions and in some cases the data derived from assessments have been used to make inferences about global constructs (e.g., Deno, 1985; Marston & Magnusson, 1985).

SUMMARY

The only structured CBA procedure with any notable empirical evaluation appears to be that of CBM (Shinn, 1989; Tindal, 1988). From a behavioral assessment perspective, the evaluative data base seems lacking in several important aspects. First, the influences of situational assessment (assessor, instructions, materials, etc.) are not well understood, especially as to how such variables may influence decisions. Recent research (Derr & Shapiro, 1989) raises serious questions about assumptions that, for example, performance on CBM probes is best conceptualized as if it were a traditional achievement test. Second, the relation of CBM measures to natural academic performances and natural environmental variables is not clear. In terms of planning classroom interventions, or of changing existing interventions, this is unsatisfactory. (The efficacy of using CBM progress monitoring to know *when* to change interventions seems supported [see Fuchs, this volume]. However, what or how to change is not necessarily derived from the use of CBM.) Evaluation of CBA/M within a behavioral assessment model would help address such concerns. Third, the use of CBM probes in improving diagnosis (i.e., easier matching of

interventions for typical problem patterns) is basically unexplored.

RECOMMENDATIONS AND FINAL CONSIDERATIONS

The continued evaluation of CBA, especially CBM, within a behavioral assessment model could address a number of intriguing and important questions. It should be acknowledged that a behavioral concept of a skill, especially in regard to basic academic skills, has not been fully explored or even well developed in a practical sense. This is important because CBM would seem to offer, if used with more direct measurements, some broad assessment of current student "skills" especially as related to the reasons that a student is not meeting naturalistic classroom expectancies (see, e.g., Lentz & Shapiro, 1986). For example, during an initial assessment of a particular student's academic problems, use of various CBA measures (including CBM probes or other curriculum-based measures) in conjunction with environmental measures could allow a decision about whether any presenting problem is related to lack of student behaviors (abilities, skills, etc.) or a failure of the academic environment to support adequate performance in required classroom/curriculum activities. Likewise, CBA/M would seem potentially useful in the analysis of variables contributing to overt classroom behavior problems. (Is the student able to access normal classroom rewards for academic performance? Is a lack of skills contributing to inappropriate behaviors?) The recommendations discussed below are intended to suggest the types of research needed to allow the fullest utilization of CBA/M in the process of solving educational problems.

Situational assessment variables and effects on CBM data. From a behavioral assessment perspective, CBM performance is not a matter of true and error components; rather, the influence of setting, assessment conditions, assessor, materials, etc., should be directly assessed. Further, these effects can and should differ across subjects. Derr and Shapiro (1989) have provided evidence that the performance of students on CBM reading probes is significantly influenced by setting, assessor, and instructions. Such influences can impact nearly all the decisions made using CBM and additional research needs to be conducted across the variety of CBM type probes, to determine how decisions may be affected.

Environmental influences on CBM performance and the relationship of CBM measures to "naturalistic" academic behaviors. Research should be

extended to examine how CBM probe performance (an analogue measure) is related to student performance on natural academic tasks, such as oral reading in reading group, seatwork across subjects, spelling across different types of written assignments, and performance on classroom tests. In some ways, this would compare two types of CBA, assessment from normal academic products, and performance on CBM probes. Gable and Hendrickson (1990) have provided a good guide for analyzing error patterns in student work in regard of identifying intervention targets. Would error patterns apparent on classwork match error patterns from CBM probes? Further, the variables that are functionally related to such performances need closer examination in order to more clearly understand how use of CBM enhances the analysis of presenting academic problems. Information from such research is required before a clear understanding of the linkage between academic assessment and intervention planning, especially in regular classrooms, is possible.

CBA/M measures and the identity of homogeneous groups of academic problems. Additional research may allow us to identify homogeneous groupings of referred children in order to maximize selection of appropriate interventions. The identification of "classes" of presenting problems that allow selection of empirically effective matching interventions is perhaps the most important goal of any diagnostic effort. For example, students with different levels of performance on CBM probes, different performance on "natural" classroom tasks, and different patterns of impinging classroom variables could be grouped conceptually and their response to different types of intervention clarified. Research results may even allow good decisions about *levels* of CBM probe performance, given types of classroom environments, that are necessary for success in regular classrooms without additional resource or "pull out" assistance. This type of research is badly needed to advance the *technology* of classroom interventions for the use of practitioners.

Generalization from academic interventions: From special to regular classrooms and within either type of classroom. Related to the research discussed immediately above is the issue of how changes in CBM measures used in progress monitoring generalize to academic behaviors in the natural classroom environment. If it is what teachers see that initiates referrals for academic assistance, then what we do about the problems must ultimately impact on such observations. Making decisions about the efficacy of academic interventions using repeated CBM measures should be examined from the "consumer" end, in terms of whether our decisions are directly related to improvement in the

behaviors about which teachers were initially disturbed. Research into this issue would involve concurrent measurement of natural classroom responding (including curriculum required daily responses) and CBM probes. Additionally, assessment of which classroom variables functionally affected this relationship would advance our understanding of generalization, and the development of generalization technology.

The stability of progress monitoring decisions. As stated, CBM has been well established as a progress monitoring system that can enhance student achievement. One problem that we have observed in our own use of CBM has been the widely different variance of individual students. Students with extreme variation on probe performance may well produce data series that result in unstable decisions about changing decisions; for example, they may require more or more frequent data points before a decision can be made about the need for change in instruction or goals. From a behavioral assessment perspective, these issues would be seen as idiosyncratic, but empirical guides for different performance patterns could be developed. Guides around number of probes across what amount of time appear to be generally lacking (see, e.g., Shinn, 1988), and such research would be useful for all users of CBM.

If districts adopt CBM procedures to replace typical evaluation procedures within the special education process there are, we believe, clear benefits. As has been concluded (e.g., Tindal, 1988), CBM appears more consistent across the wide range of necessary decisions, use of CBM in progress monitoring appears to enhance achievement (Fuchs, 1989), and CBM may improve program evaluation in special education (e.g., Tindal, 1989). If professionals such as school psychologists adopt CBM and other CBA procedures during academic assessment we also believe that children would benefit and we have suggested research to enhance the validity of decisions made in such assessments. However, if the traditional refer-test-place procedure remains virtually intact and CBM data replaces other "gatekeeping" data, then there may be little effect on children outside of special education, and only then to the extent that structured progress monitoring occurs. Although continued CBA research within the placement process, especially regarding decision stability would be helpful, research into CBA/M from a behavioral assessment perspective would greatly enhance intervention assistance efforts for all "at risk" students. Finally, such research would also illuminate the efficacious selection of interventions within special education programs.

The goals and objectives established for this paper were clearly stated at the outset. In our examination of academic skill assessment it

has been argued that the behavioral model is most appropriate for use understanding functional relationships between assessment data and environmental conditions. The discussion suggests that CBM procedures have often been used and interpreted within a traditional measurement model, although other research more consistent with the logic expressed herein has begun to appear. Although direct observation and measurement of classroom behavior is expensive, we argue that measurement of natural classroom events are the standard against which less direct measures (e.g., CBM probes) be evaluated. There is much to be learned about the relationship between performance in the natural context in which academic performance occurs and CBM data.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

American Psychological Association. (1981). Ethical principals of psychologists. *American Psychologist, 36,* 633-638.

Arter, J. A., & Jenkins, J. R. (1979). Differential diagnosis--prescriptive teaching: A critical appraisal. *Review of Educational Research, 49,* 517-555.

Benes, K. M., & Kramer, J. J. (1989). The behavioral tradition in schools (and miles to go before we sleep). In J. Hughes & R. Hall (Eds.), *Handbook of cognitive therapy approaches in schools* (pp. 15-36). New York: Guilford.

Bickel, W. E., & Bickel, D. D. (1986). Effective schools, classrooms, and instruction: Implications for special education. *Exceptional Children, 52,* 489-500.

Blankenship, C. S. (1985). Using curriculum-based assessment data to make instructional decisions. *Exceptional Children, 52,* 233-238.

Cone, J. (1981). Psychometric considerations. In M. Hersen & A. Bellack (Eds.), *Behavioral assessment: A practical handbook* (2nd ed., pp. 38-70). Elmsford, NY: Pergamon Press.

Conoley, J. C., & Gutkin, T. B. (1986). School psychology: A reconceptualization of service delivery realities. In S. N. Elliott & J. C. Witt (Eds.), *The delivery of psychological services in schools: Concepts, processes, and issues* (pp. 393-424). Hillsdale, NJ: Erlbaum & Associates.

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

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

Deno, S. L., & Mirkin, P. K. (1977). *Data based program modification: A manual.* Reston, VA: The Council for Exceptional Children.

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

Derr, T. F., & Shapiro, E. S. (1989) A behavioral evaluation of curriculum-based assessment. *Journal of Psychoeducational Assessment, 7*, 148-160.

Fuchs, L. S. (1989). Evaluating solutions: Monitoring progress and revising intervention plans. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 153-181). New York: Guilford.

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

Fuchs, L. S., Tindal, G., & Deno, S. L. (1984). Methodological issues in curriculum-based reading assessment. *Diagnostique, 9*, 191-207.

Gable, R. A., & Hendrickson, J. M. (Eds.). (1990). *Assessing students with special needs: A sourcebook for analyzing and correcting errors in academics.* NY: Longman.

Gallessich, J. (1974). Training the school psychologist for consultation. *Journal of School Psychology, 12*, 138-149.

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

Goldwasser, E., Meyers, I., Christenson, W., & Graden, J. (1983). The impact of P. L. 94-1421 on the practice of school psychology: A national survey. *Psychology in the Schools, 20*, 153-165.

Greer, R. D. (1983). Contingencies of the science and technology of teaching and prebehavioristic research practices in education. *Educational Researcher, 12*(1), 3-9.

Haring, N., & Eaton, M. (1978). Systematic instructional procedures. An instructional hierarchy. In N. Haring, T. Lovitt, M. Eaton, & C. Hansen (Eds.), *The fourth R: Research in the classroom* (pp. 23-41). Columbus, OH: Charles E. Merrill.

Hartmann, D. P., Roper, B. L., & Bradford, D. C. (1979). Some relationships between behavioral and traditional assessment. *Journal of Behavioral Assessment, 1*, 3-21.

Haynes, S. N., & Wilson, C. C. (1979). *Behavioral assessment.* San Francisco: Jossey-Bass.

Hersen, M., & Bellack, A. S. (Eds.). (1981). *Behavioral assessment: A practical handbook* (2nd ed.). Elmsford, NY: Pergamon Press.

Hops, H. (1971). The school psychologist as behavior manager consultant in a special class setting. *Journal of School Psychology, 9*, 473-483.

Kazdin, A. (1984). *Behavior modification in applied settings* (3rd ed.). Homewood, IL: Dorsey Press.

Kramer, J. J., Henning-Stout, M., Ullman, D. P., & Schellenberg, R. P. (1987). The viability of scatter analysis on the WISC-R and SBIS: Examining a vestige. *Journal of Psychoeducational Assessment, 5*, 37-47.

Kramer, J. J., & Peters, G. J. (1986). Service delivery to rural schools: Conceptual and logical hurdles. In S. N. Elliott & J. C. Witt (Eds.), *The delivery of psychological services in schools: Concepts, processes, and issues* (pp. 202-225). Hillsdale, NJ: Erlbaum & Associates.

Lentz, F. E. (1988). Direct observation and measurement of academic skills: A conceptual review. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in the schools: Conceptual foundations and practical applications* (pp. 76-120). New York: Guilford Press.

Lentz, F. E., & Shapiro, E. S. (1986). Functional assessment of the academic environment. *School Psychology Review, 15*, 346-357.

MacMahen, G., & Barnett, D. (1985). Discrepancy score analysis: A computer simulation of classification stability. *Journal of Psychoeducational Assessment, 4*, 363-375.

Marston, D., & Deno, S. L. (1981). *The reliability of simple, direct measures of written expression* (Research Report No. 50). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities. (ERIC Document No. 212 663)

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

Marston, D., & Magnusson, D. (1988). Curriculum based measurement: District level implementation. In J. Grader, J. Zins, & M. Curtis (Eds.), *Alternative educational delivery systems: Enhancing options for all students* (pp. 137-172). Washington, DC: National Association of School Psychologists.

Martens, B. K., & Witt, J. C. (1988). Ecological behavior analysis. In I. M. Hersen, R. M. Eider, & J. P. M. Miller (Eds.), *Progress in behavior modification* (Vol. 22, pp. 115-140). Beverly Hills, CA: Sage.

Nelson, R. O., & Hayes, S. C. (1981). Nature of behavioral assessment. In M. Hersen & A. Bellack (Eds.), *Behavioral assessment: A practical handbook* (2nd ed., pp. 3-37). Elmsford, NY: Pergamon Press.

Shapiro, E. S. (1990). An integrated model for curriculum-based

assessment. *School Psychology Review, 19,* 331-349.

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. (1989). Identifying and defining academic problems: CBM screening and eligibility procedures. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 90-129). New York: Guilford.

Shinn, M. R., & Marston, D. (1985). Differentiating mildly handicapped, low achieving and regular education students: A curriculum-based approach. *Remedial and Special Education, 6,* 31-45.

Tindal, G. (1988). Curriculum based measurement. In J. Graden, J. Zins, & M. Curtis, (Eds.), *Alternative educational delivery systems: Enhancing options for all students* (pp. 111-135). Washington, DC: National Association of School Psychologists.

Tindal, G., Fuchs, L. S., Fuchs, D., Shinn, M. R., Deno, S. L., & Germann, G. (1985). Empirical validation of criterion-referenced tests. *Journal of Educational Research, 78,* 203-209.

Tindal, G. (1989). Evaluating the effectiveness of educational programs at the systems level using curriculum-based measurement. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 202-238). New York: Guilford.

Tucker, J. A. (1985). Curriculum-based assessment. *Exceptional children* (Special Issue), *52* (3).

White, O., & Liberty, K. (1976). Behavioral assessment and precise educational measurement. In N. Haring & R. Schliefelbusch (Eds.), *Teaching special children* (pp. 31-69). New York: McGraw Hill.

Witt, J. C., & Gresham, F. M. (1985). Review of the Wechsler Intelligence Scale for Children-Revised. In J. V. Mitchell, Jr. (Ed.), *The ninth mental measurements yearbook* (pp. 1716-1719). Lincoln, NE: Buros Institute of Mental Measurements.

5

Curriculum-Based Assessment: Implications for Psychoeducational Practice

Edward S. Shapiro

Lehigh University

The topic of this chapter places me somewhere between Carnac the Magnificent and a crystal ball gazer! On the one hand, I am being asked to look into the future and discuss the potential implications of curriculum-based assessment (CBA) for psychoeducational practice. Although my graduate students believe I may have superhuman powers and can be all places at the same time, fortune telling was never one of my talents. On the other hand, like Carnac, I obviously believe that CBA is an answer, but I'm not sure what the questions are going to be. In this paper I assume that all questions asked have the same answer: "Use CBA."

When a district decides to adopt CBA as a measurement procedure, impacts are anticipated on the service delivery method, accountability procedures, and role functions within that district. The way in which CBA is adopted, the particular model of CBA employed, and the acceptance of CBA in the district will all play a part in the degree to which each of these aspects of the district are affected.

Implementing CBA districtwide obviously will have implications that may alter the entire system. Equal impact may be noted when CBA is implemented on an individual basis. A single teacher may choose to use CBA within his or her classroom. A single psychologist may choose

to use CBA as a means to enhance service delivery. A single resource room teacher may choose to implement CBA for a particular class. Further, the ways in which CBA are used may not be individualized. A single teacher may choose to provide progress monitoring on long-term goals. A resource room teacher may choose to implement progress monitoring for long-term goals and write IEP objectives using CBA. A psychologist may choose CBA as a mechanism for conducting initial evaluations and recommending intervention strategies.

Use of CBA by individuals has implications that are somewhat different than when CBA is used in an entire system. For example, when an individual uses CBA to make eligibility decisions, one obviously cannot use CBA alone but must find a way to integrate CBA and traditional assessments. Additionally, using CBA to identify targets for intervention can be valuable only if the delivery system supports intervention planning rather than educational diagnostic decision making.

Recognizing that there are some differences between using CBA with an individual versus large-scale application, I will confine my comments to the implications of CBA when employed on a large-scale, districtwide basis.

IMPLICATIONS FOR SERVICE DELIVERY

How Should Eligibility for Special Education Be Determined? Use CBA

Certainly not the intention of developers of CBA, much attention has been given to its potential use as a mechanism to determine the eligibility of students for classes for the mildly handicapped. This has been particularly true of the curriculum-based measurement (CBM) model of CBA. From the onset of the dissemination of this model, researchers published many studies that examined the concurrent and criterion-related validity of CBM. These studies typically would determine the degree to which already identified groups of learning-disabled (LD) and non-LD students would be differentiated by CBM measures (e.g., Deno, Marston, Shinn, & Tindal, 1983; Deno, Mirkin, & Chiang, 1982; Marston & Deno, 1982; Shinn, Ysseldyke, Deno, & Tindal, 1986). These studies showed that CBM measures could distinguish between already classified learning-disabled, non-learning-disabled, and Chapter I students (Marston, Tindal, & Deno, 1984; Marston, Mirkin, & Deno, 1984). Further studies addressing the criticism of using

intact groups reported that CBM measures "predict correct membership in special education about as accurately as the commercial measures of achievement" (Tindal, 1988).

Using CBA, and CBM in particular, as a mechanism to determine eligibility for special education appears to have some research support. By employing ratios of expected to actual performance, called discrepancy ratios, a ratio of 2.0 to 2.5 appears to result in the equivalent percentage of students being classified as eligible for special education as traditional methods. This was true of most grades except first and second, where such a ratio resulted in a significantly higher percentage of students identified as handicapped (Marston, Tindal, & Deno, 1984). What are the implications for service delivery of putting such a system in place? What are the potential impacts on individual students when their eligibility for special education has been based on CBA?

By using CBA- or CBM-type measures to determine eligibility for special education, the criteria for entering special education become clearly demarcated. The degree to which students must fall behind to be eligible is empirically determined and is based on observable student performance of required tasks, rather than some unobservable, mystical entity entitled potential. Empirically based criteria for determining special education eligibility, particularly learning disabilities, would be a welcome relief from the way these decisions currently are being made. Indeed, the Panel on Selection and Placement of Students in Programs for the Mentally Retarded (Heller, Holtzman, & Messick, 1982) raised serious questions about the use of traditional measurement procedures (e.g., IQ tests, standardized achievement tests) in the decision to declare students eligible for special education services.

To effectively implement a CBA-based eligibility decision-making model, local norms must be developed. Although there is little research into parameters of the norming process for CBM (e.g., extent of population needed to be sampled, using building versus districtwide norms, how to handle the problem of multiple basal reading series used within the same district), the time, energy, and expense of collecting and developing local norms must be recognized. In some of the norming projects I have been aware of in Iowa and Pennsylvania, the cost of collecting norms has been borne by grants from states or local districts. Although this is appropriate for pilot projects, there must be mechanisms built into systems to perpetuate the collection of norms. Without this perpetual motion, it is unlikely that ongoing updating of local norms will occur.

Another implication of using CBA for eligibility decision making is related to establishing criteria for exiting special education. It seems logical that CBA can be used as much to enter students into special education as it can to establish criteria for exiting. One of the most significant problems facing special education is that once students have been declared eligible, they rarely move out. Declassification statistics are difficult to find; however, most school professionals will tell you that most students carry their special education label with them for the duration of their school careers. By using CBA, one could identify the level of performance equivalent to, for example, the lowest reading group or math group in an elementary school. When such a level is established by the child receiving special education services, and maintained for a specified period of time within a regular education setting, the student may be declassified as needing special education. Clearly, this should alter the rates of entrance and exit from special education.

Cone (1988) has described a behavioral assessment procedure called template matching that could be very valuable for using CBA to determine exit criteria from special education. In template matching, target behaviors are identified and assessed on those judged to be "average" responders. The ranges of these behaviors across students are graphed using box and whisker plots. Behaviors of problematic youngsters are assessed to determine how their levels of the identical behaviors match the nonproblematic students. Hoier, McConnell, and Pallay (1987) presented an excellent example of template matching in the evaluation of handicapped preschool children. In their study, they identified which behaviors would be problematic for children moving from preschool to kindergarten and kindergarten to first grade. Hoier et al. did not go the additional step of deriving intervention strategies to teach these skills, but the template matching procedure was an excellent way to show clearly which behavior patterns may be problematic when handicapped students are mainstreamed.

A similar procedure could be employed using CBA. Data collected from nonhandicapped "average" peers may offer the template and targets for interventions among handicapped youth. Indeed, this is often the case when IEP goals are set and could be used to set exit criteria as well. Further, using this strategy in the assessment of the academic ecology could also lead to targets for intervention that may need to be addressed, in order to have the student attain success in the regular education setting.

The use of CBA as a decision-making model for special education eligibility clearly requires some policy changes. Policy at state levels

must support the opportunities for local districts to experiment and then permanently replace existing models of decision making. Support is not always easy to come by, although large districts such as Minneapolis, as well as Departments of Education like Iowa, have been able to solicit support. In particular, there are always concerns raised about ignoring the *potential* part of the equation in identifying learning-disabled students. Alteration of this part of policy requires changes in basic assumptions about predicting success in school. As articulated by Marston and Magnusson (1988), the best predictor of reading performance cannot be the degree to which a student answers questions about history, does puzzles, and copies designs.

At both the district and building levels, there are needs for understanding and accepting CBA as a viable alternative to current ways of making decisions about student performance. Principals, teachers, and district administrators must be convinced that the measurement systems advocated by CBA have the conviction of more traditional approaches. They must be convinced that their decisions indeed are supported by teachers, parents, and state departments of education. At present, little is known about the acceptability of CBA as viewed by various education professionals. In a pilot study among two samples of teachers, Turco and I (1988) found that CBA does indeed show significantly higher levels of acceptance as rated on a measure of assessment acceptability. In contrast, among a nationally sampled group of school psychologists in the same study, no differences are evident in acceptance of CBA compared to traditional achievement measures. When teachers and psychologists are compared, however, there does appear to be a significantly higher acceptance rating of CBA by teachers compared to school psychologists. Although I stress the preliminary nature of these findings, both the development of an assessment acceptability scale and the initial findings of teachers having higher acceptability of CBA than psychologists begin to point out some of the issues that must be faced, in order to reach the acceptance level where CBA may impact successfully upon a system.

One important problem raised by using CBA as a means of deciding eligibility for special education services is the political reality of advocacy groups. Many administrators willing to consider CBA must also consider the impact on numbers of students declared eligible. Altering the discrepancy ratio empirically alters those who are eligible to receive services. Fears of this nature drive advocacy groups into a frenzy. I have seen firsthand the rejection of excellent and innovative ideas that had the support of teachers and administrators because of fears of advocacy group reaction.

Another consideration in using CBA as a districtwide measurement procedure relates to the consistency of curriculum employed across the district. For example, there may be problems related to particular basal reading series across the district. CBA results may not be easily generalized across curricular series. In some districts where the selection of basal reading series are not standardized across schools, this can present significant problems. Additionally, if students within special education classes are judged on different curricula than those in regular education, there may be difficulties in trying to make effective comparisons and decisions about how special education students would be doing if they were being instructed within the regular education environment.

In general, the implementation of a CBA model for declaring students eligible for special education solves some problems and creates new ones. Decisions using CBA may be viewed as potentially less susceptible to racial and ethnic biases (Shinn & Tindal, 1988), often considered significant problems in the use of standardized tests with children of minority groups. While it is true that CBA does not bring with it the content validity problems of racial bias evident on some standardized tests, it may not change the overrepresentation issue of minorities in special education. More research clearly is needed to confirm this, but it seems that CBA could be as biased as the curriculum, if you define bias in terms of the percentages of assessed students found eligible for special education.

CBA may also address the question of subjectivity in decision making. Students declared eligible are done so based on empirical findings, and decisions regarding one's sense that a student is learning disabled, for example, are less likely to occur. Decision-making biases of multidisciplinary teams, as found by Ysseldyke and colleagues (e.g., Algozzine & Ysseldyke, 1981; Epps, Ysseldyke, & McGue, 1984), should be limited, although their findings have not been consistently replicated (Huebner, 1987; Huebner & Cummings, 1985).

Successfully solving some problems, CBA-based eligibility decisions introduce other serious problems. How does a district set its discrepancy ratio to determine eligibility? One can envision a district being told that its special education budget was just cut by 10%. A quick accounting of costs may show that the district can meet its budgetary constraints if it changes its discrepancy ratio from 2.0 to 2.5. Indeed, in one district I am aware of, the district superintendent decided that the percentage of special education students in their district would be no more than 3.0% of the district population. To accomplish this goal, a discrepancy ratio was altered. This type of problem and solution can

create significant discrepancies in who does and does not receive special education services. As such, the decision of who is served is based on politics and not need, potentially raising serious legal as well ethical concerns.

Unless a district makes a substantial and long-term commitment to the development of norms, supports those who are assigned to collect data, supports the maintenance of the data base, and provides ongoing training as staff in the district changes, the success of using CBA to make eligibility decisions is questionable. Further, if this is the only way in which CBA is employed in a district, one legitimately should question its cost-effectiveness. Making an argument for the cost-effectiveness of systemwide implementation of CBA requires use of the data for more than special education eligibility decision making.

How do I design effective interventions for classroom problems?
Use CBA

Advocates of CBA consistently suggest that the primary value of CBA procedures is the ability to use these procedures to identify effective intervention strategies for academic problems. The evaluation of variables related to the instructional ecology (Lentz & Shapiro, 1986; Ysseldyke & Christenson, 1987), combined with the assessment of individual skills, provides a framework for suggesting potential strategies that may be effective in remediating and preventing academic difficulties. Recommended strategies for intervention usually are based only partially on the data obtained during the assessment. These data offer "educated guesses" as to what may be an effective procedure. However, the choice of appropriate interventions may just as well be based on the combined knowledge, experience, and preference of the teacher, psychologist, or other educational consultant. Some intervention procedures, like classwide peer tutoring, are not really derived as strategies likely to be effective based on the data alone, but are employed as overall instructional strategies because of their proven effectiveness.

There are several models of CBA that do focus explicitly on the development of intervention strategies. Curriculum-based evaluation (CBE), developed by Howell and Morehead (1987), uses a task analysis approach to examine errors in academic responding and then designs instructional programs to teach the needed components or subcomponents of skills. Likewise, Blankenship (1985) and Idol, Nevin, and Paolucci-Whitcomb (1986) proposed a model of CBA that relies heavily on evaluating acquisition of specific curriculum objectives. Perhaps the model with the most substantial link to designing intervention strategies is that developed by Gickling and colleagues

(Gickling & Havertape, 1981; Gickling & Thompson, 1985). Their model is based on the assessment of known and unknown material a student is being taught, followed by the teaching of unknown material under specified ratios to assure student success.

There are several potential implications in using CBA to derive intervention strategies. First, an underlying assumption of CBA is that the academic deficiencies evident in the classroom are the result of an interaction between the instructional ecology and individual student skill mastery. Learning does not occur in a vacuum but in the context of a teaching environment. This component in the learning equation cannot be ignored. Traditional assessment and intervention strategies are often focused solely on the individual. Rarely is the instructional environment considered as the cause of the student's problems. When Johnny cannot spell, it is because he cannot phonetically analyze the words. When he cannot add, it is because he has difficulties in mental operations. How often does the teacher conclude that Johnny cannot spell because corrective feedback occurs too infrequently? Or that he cannot add because the contingencies for performance are not sufficient?

Using CBA to derive intervention strategies requires a shift from viewing problems as person oriented to person/environmental interactions. This shift is more easily said than done. All of us have had numerous experiences with school personnel of all types, including teachers, psychologists, etc., where the inferred cause of identified problems is quickly decided to be skill and personal deficiencies in the student (e.g., auditory perception, dependent personality). Shifting to a person/environment interactional framework will not be accepted easily because it requires evaluation of instruction and instructional components and, by implication, people's ability to teach. This approach to assessment is uncommon and may have limited acceptability among the consumers of this information (i.e., teachers, parents).

A second implication of using CBA for intervention planning is the increased pressure to move a district toward preplacement or prereferral service delivery. There has been significant movement in this direction across the country. Using CBA within a prereferral service delivery model will require more than the typical way in which child-study teams are conducted. The team must have a mechanism to respond to the data collection process. Many child-study teams focus upon determining if students are eligible for special education. This is accomplished by having each member of the team report the results of his or her assessment, with the team jointly deciding if the data suggest the student meets the eligibility criteria. Using CBA within a child-

study team process cannot be simply a reporting of what each member of the team found. Clearly, child-study teams need to learn how to use CBA data to make intervention decisions. They need to learn how to report effectively CBA data beyond consideration of eligibility of services. To implement a service delivery system of this type requires enormous retraining and rethinking of how services are being provided.

A third implication of using CBA data in intervention planning can be seen in the increased instructional decision-making capabilities granted to teachers by this approach. Witt and Martens (1988), among others, suggest strongly that teacher empowerment is critical for successful implementation of any alternative service delivery model. Using CBA for planning interventions offers teachers the perfect opportunity for their expertise to become a critical element in choosing intervention strategies. How comfortable are school administrators with the added power teachers attain when they are permitted to be responsive to their own data collection process that comes with this model? How comfortable are school psychologists in trusting the judgments of teachers?

Another implication of using CBA data to plan interventions is that it may help to remove the mystique of the testing process. Many times, school personnel seem to regard test results, particularly group test results, as the only legitimate means of answering questions regarding student outcome and program success. How many times have psychologists been asked the question, "So what's his IQ?"? When the scores are reported reluctantly, the response is, "No wonder he's having trouble." The IQ score is viewed as some mystical number that identifies, explains, and permits the failure of some students. In contrast, by using CBA data, the performance of the student becomes the criteria for making decisions. There are no mystical concepts or hidden messages. What you see is what you get! When the student is not doing well, we know it, we can see it, we can empirically verify it, and we can ask the question of how to change it! This concept may not be acceptable to many individuals who are trained in models stressing the importance of underlying psychological and educational processes.

Another potential implication of using CBA to assist in intervention selection assumes that teachers and other educational consultants have knowledge of effective intervention strategies. It has been my experience, in three states where CBA has been implemented, that one should not expect teachers to know how to use procedures like peer tutoring, cooperative learning, learning strategies, self-management, or effective use of contingency management. One way to facilitate the selection of intervention strategies is to consider the collective wisdom

of many teachers and education professionals. Using teacher assistance teams or similar concepts has often been successful and does not assume any one individual has the answers. This has been reported in the literature where CBA has been used to assist decision making (Fuchs & Fuchs, 1988; Marston & Magnusson, 1988). Obviously, the provision of training, whereby effective interventions may be taught to teachers and educational consultants, is needed.

One potential concern about moving towards CBA as a critical component of service delivery is the possibility that the use of prereferral intervention models may simply delay, rather then prevent, the placement of students into special education. Clearly, carefully controlled longitudinal research studies, examining the impact of prereferral intervention models when employed systemwide, are needed to determine whether this is occurring.

ACOUNTABILITY

How should I write IEP goals? Use CBA

Using CBA to write IEP goals probably represents one of the most important ways that CBA can be employed. When CBA is used to write IEP goals, we are suddenly thrust into an age of accountability. Teachers can determine objectively if their students meet goals set earlier in the year. Students can see, monitor, and evaluate their own progress toward goals. In fact, students can help write these goals. How often do students attend and contribute to their own IEP goals? How often do students know if they are making progress towards these goals?

Additionally, parents can be offered concrete evidence of educational change. For example, a parent of a boy came to me for an evaluation. The boy had been in a self-contained classroom for students with learning disabilities for 2 years. The mother expressed concern about her son's academic progress after finding he made no improvement for 2 years on the results of the California Achievement Tests. Questioning the value of the placement in a class for students with learning disabilities, she asked for an opinion regarding how much progress he actually had made. An evaluation using teacher interviews, direct observation of the instructional environment, and CBA found he was indeed making significant progress and mastery of skills. Further, the classroom structure employed in his self-contained setting was excellent, and embodied most of the critical variables of effective teaching. After recommending that IEP goals be rewritten in CBA terms, along with progress monitoring, the mother, teacher, and student began to see his

rate of progress within 4 weeks of starting monitoring.

An implication of this increased accountability is the potential misuse of these data. CBA data should not be used alone as indicators of successful teaching. Just because some students do not meet their goals does not imply that the teacher is a poor teacher. Indeed, what seems to be critical is that decisions regarding outcomes of instructional interventions are being evaluated empirically. Fuchs, Deno, and Mirkin (1984) demonstrated that simply getting teachers to use frequent progress monitoring may result in substantial improvements in academic achievement. In many ways, the use of progress monitoring may provide the evidence needed to evaluate clearly the potential necessity for a more restrictive educational placement. For example, if a teacher showed data that suggested a series of unsuccessful interventions were tried during data collection, this may be strong evidence that although the student failed to progress, the teacher indeed was responsive to the data collection process. In contrast, without the collection of these data, teachers may be viewed as failures, based solely on the lack of progress of their students. If this happens, CBA would never gain the sanction of teacher unions!

A related accountability issue is the ability for CBA to reflect programmatic success. By aggregating data across individuals, one is able to obtain a concrete picture of the success of an entire educational program, such as a special education resource room program. For example, Marston and Magnusson (1988) described how CBM was employed districtwide, infused into the screening, identification, instructional planning, monitoring, and evaluation of services for students exhibiting academic skills problems.

Role Functions

Putting CBA in place will alter significantly the roles of several persons typically involved in the assessment process. Special and regular education teachers are suddenly thrust into a very important and critical role in the multidisciplinary team. These persons become crucial points of information and consultation. They are no longer regarded as simply making referrals to professionals for advice and consultation. Instead, they are viewed as key components in the assessment and remediation process.

Resource room teachers may play a particularly important new role. In most districts, resource room teachers serve in a direct service capacity. They often have their own room where students come for remediation for a portion of the day. Rarely are the knowledge and

skills of these persons made available on a regular basis to teachers in regular education classrooms. Yet, as the service delivery method shifts toward using CBA, these persons can play a crucial role in enhancing the consultation process. Further, these persons can assist in the preventative nature of such services.

School psychologists often have significant adjustments to make. These professionals may feel that CBA does not belong in the realm of their profession. They may see it as strictly belonging to the teachers and therefore, psychologists may reject CBA as not vital to their assessment. This is problematic, since in many districts where CBA has been implemented, it has been the school psychologists who have been instrumental in leading the charge toward its acceptance. School psychologists working in districts where CBA is being employed, particularly as a prereferral model, must examine their current methods of service delivery and recognize the potential of their contribution to the team. Indeed, school psychologists are often some of the most knowledgeable persons in the district on collecting local norms, on the psychometric properties of measurement, and on analyzing and interpreting data. Using school psychologists in this way could broaden their roles far beyond their traditional service delivery model.

Administrators also may see the benefits of using CBA. One of the most common issues raised by administrators is the need to design effective schools. Such schools contain an atmosphere of support, collaboration, collegiality, and professionalism. It seems that providing an empirically based mechanism to evaluate instructional decisions for students could assist administrators in providing valuable feedback to teachers and parents. Likewise, administrators are always faced with the need to allocate carefully their shrinking resources. At times, the allocation of these resources can be difficult, raising questions about administrative priorities. Administrators may be forced to overallocate resources to certain students simply because they fit into a particular category of disability. Yet, these same students may not need the equivalent level of resources as other students who also meet such categorical classifications. CBA offers a potential mechanism to determine instructional requirements based on educational needs rather than category. As such, decisions regarding the allocation of resources can be made based on empirical data and potentially offer more equitable distributions of available support. Ultimately, accountability for these decisions is also provided as data continue to be collected.

CONCLUSIONS AND FINAL REMARKS

Clearly, there are numerous ways in which CBA would impact upon every aspect of a school district. Individual students, regular education teachers, special education teachers, administrators, school psychologists, educational consultants, and others typically involved with the multidisciplinary team all will be affected by the full implementation of CBA.

Let me return to the reality of the situation and leave fortune telling behind. There have been a few successful demonstrations reported of wide-scale attempts to use CBA as a mechanism for altering service delivery. In particular, the efforts in Minneapolis (Marston & Magnusson, 1988), Pine County Cooperative School District (Germann & Tindal, 1985), and the State of Iowa (Grimes & Reschly, 1986) stand out. In both cases, the impact upon the system was clear. Marston and Magnusson (1988) indicated the role of the resource room teacher has changed dramatically to include increased expectations of individual students, along with increased accountability for the effectiveness of instruction. Essentially, using CBA as a model for evaluation and designing instruction resulted in expectations of behavior change, and directly implied that teachers can be instrumental in altering student performance. Likewise, this expectation led to resource room teachers feeling more accountable for their instruction.

The role of the school psychologists also changed in Minneapolis and Iowa. Instead of the traditional responsibilities of performing evaluations for eligibility, school psychologists were assigned full-time responsibilities to organize and oversee the implementation of CBM. This included coordinating the data collection and norming process, and providing in-service, data analysis, and other activities in support of the program. Interestingly, because the Minneapolis system employed resource room teachers as the primary "doers" of CBM, school psychologists were still expected to maintain responsibilities in consultation, and to direct treatment evident prior to the CBM program. However, Marston and Magnusson (1988) noted that a report provided by Canter (1986) showed psychologists to be spending proportionally more time on fewer cases, while increasing the amount of time spent in consultation.

Administrators' roles have also been altered by the implementation of the Minneapolis CBM project. With the objective measurement provided by CBM, resource allocation, training needs, and policy development are consistent with a system based on student outcomes. Noting needs for in-service training, as well as determining needed

teaching resources, can be determined by looking directly at school-based performance levels.

Despite the many positive and valuable outcomes possible with systemwide implementation of CBA, there are obviously important concerns and considerations. Probably one of the most critical elements for the effective implementation of CBA is acceptance of the assumptions, methodology, and value of such a measurement system. Without a change in these basic attitudes, CBA is likely to be another passing educational promise that will never reach its potential.

REFERENCES

Algozzine, B., & Ysseldyke, J. E. (1981). Special education services for normal children: Better safe than sorry? *Exceptional Children, 48*, 238-243.

Blankenship, C. S. (1985). Using curriculum-based assessment data to make instructional decisions. *Exceptional Children, 52*, 233-238.

Canter, A. (1986). *1986 time/task study*. Minneapolis, MN: Minneapolis Public Schools, Department of Special Education, Psychological Services.

Cone, J. D. (1988). Psychometric considerations and multiple models of behavioral assessment. In A.S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (pp. 42-66). New York: Pergamon.

Deno, S. L., Marston, D., Shinn, M. R., & Tindal, G. (1983). Oral reading fluency: A simple datum for scaling reading disability. *Topics and Teaching in Learning Disabilities, 2*(4), 53-59.

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

Epps, S., Ysseldyke, J. E., & McGue, M. (1984). "I know one when I see one"—Differentiating LD and non-LD students. *Exceptional Children, 7*, 89-101.

Fuchs, D., & Fuchs, L. S. (1988). Mainstream assistance teams to accommodate difficult-to-teach students in general education. In J.L. Graden, J.E. Zins, & M.J. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 49-70). Washington, DC: National Association of School Psychologists.

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*, 449-460.

Germann, G., & Tindal, G. (1985). An application of curriculum-based assessment: The use of direct and repeated measurement. *Exceptional Children, 42*, 244-265.

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

Gickling, E. E., & Thompson, V. P. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52*, 205-218.

Grimes, J., & Reschly, D. (1986). *Relevant education assessment and intervention model: Project RE-AIM*. Des Moines, IA: Iowa Department of Education.

Heller, K. A., Holtzman, W. H., & Messick, S. (Eds.). (1982). *Placing children in special education: A strategy for equity*. Washington, DC: National Academy Press.

Hoier, T. S., McConnell, S., & Pallay, A. G. (1987). Observational assessment for planning and evaluating educational transitions: An initial analysis of template matching. *Behavioral Assessment, 9*, 6-20.

Howell, K. W., & Morehead, M. K. (1987). *Curriculum-based evaluation for special and remedial education*. Columbus, OH: Merrill.

Huebner, E. S. (1987). The effects of type of referral information and test data on psychoeducational decisions. *School Psychology Review, 16*, 382-390.

Huebner, E. S., & Cummings, J. A. (1985). The impact of sociocultural background and assessment data upon school psychologists' decisions. *Journal of School Psychology, 23*, 157-166.

Idol, L., Nevin, A., & Paolucci-Whitcomb, P. (1986). *Models of curriculum-based assessment*. Rockville, MD: Aspen.

Lentz, F. E., & Shapiro, E. S. (1986). Functional assessment of the academic environment. *School Psychology Review, 15*, 346-357.

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

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

Marston, D., Mirkin, P. K., & Deno, S. L. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *Journal of Special Education, 18*, 109-118.

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

Shinn, M. R., & Tindal, G. (1988). Using student performance data in academics: A pragmatic and defensible approach to non-discriminatory assessment. In R.L. Jones (Ed.), *Psychoeducational assessment of minority group children* (pp. 383-410). Berkeley, CA: Cobb & Henry.

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

Tindal, G. (1988). Curriculum-based measurement. In J.L. Graden, J.E. Zins, & M.J. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 111-136). Washington, DC: National Association of School Psychologists.

Turco, T. L., & Shapiro, E. S. (1988). Teacher and psychologist rated acceptability of traditional and curriculum-based assessment. Unpublished manuscript, Lehigh University, Bethlehem, PA.

Witt, J. C., & Martens, B. K. (1988). Problems with problem-solving consultation: A re-analysis of assumptions, methods, and goals. *School Psychology Review, 17*, 211-226.

Ysseldyke, J. E., & Christenson, S. L. (1987). *The Instructional Environments Scale*. Austin, TX: PRO-ED.

CBA: An Assessment of Its Current Status and Prognosis for Its Future

Mark R. Shinn and Roland H. Good, III

University of Oregon

The very fact that curriculum-based assessment (CBA) forms the basis of a topic-driven conference at the center of American educational and psychological measurement (i.e., the Buros Institute) is testimony that the strategies are receiving a substantial amount of professional attention. Although debate continues regarding to whom and when the term curriculum-based assessment should be ascribed (Coulter, 1988), without question, its prominence has grown considerably in the last 10 years. Within the last 5 years, school psychology and special education have seen their flagship journals, *School Psychology Review* and *Exceptional Children*, devote special volumes to CBA. National organizations such as the National Association of School Psychologists (NASP) and the National Coalition of Advocates for Children (NCAS) have encouraged the use of CBA for decision making with handicapped students (NASP/

NCAS, 1985). Interpretations of recent litigation also have been construed to suggest use of CBA strategies (Reschly, Kicklighter, & McKee, 1988a; Reschly, Kicklighter, & McKee, 1988b; Reschly, Kicklighter, & McKee, 1988c).

This chapter seeks to examine CBA's future as an assessment strategy from a perspective of school systems change (Sarason, 1982) and adoption of technological innovations (Rogers, 1983). To understand the school-change process, Hall and Hord (1984) maintain that change agents must consider the perspective of the *implementors* of the innovation. Using what they call a Concerns-Based Adoption Model (CBAM), Hall and Hord (1984) propose that implementors' concerns about change progress through a sequence of seven stages: (a) awareness, (b) informational, (c) personal, (d) management, (e) consequence, (f) collaboration, and (g) refocusing. An individual's concerns about innovation are not confined to any one stage, however. The seven concerns are divided into four general categories. Awareness is categorized as an *unrelated concern*, where the implementor generally is only somewhat cognizant of the innovation. Informational and personal concerns are *self concerns*, where the implementors' reactions are centered primarily on how the innovation affects them. Management is a *task concern*, where consideration is given to how best to use the innovation. Consequence, collaboration, and refocusing are *impact concerns*, where attention is shifted to the potential effects of the innovation on clients.

Each stage of Hall and Hord's CBAM model requires a different approach to influencing and facilitating the change process. At best, we believe the field of education, and more specifically special education and school psychology, is currently at the awareness and informational stages with respect to the implementation of CBA. Professionals are being exposed to CBA and are gathering information. We believe that an analysis of the future of CBA will require us to examine first the extent of professionals' knowledge regarding CBA. We will accomplish this task in two ways. First, we will identify briefly the major innovators in CBA and where their information is being disseminated. Second, we will analyze the major critiques of CBA (Lentz & Shapiro, 1986; Lombard, 1988a; Lombard, 1988b; Taylor, Willits, & Richards, 1988) under the premise that one gains an understanding of *what* is being communicated by *how accurately* it is described by others than the innovators themselves.

Before we can consider widespread adoption of CBA procedures, we must move beyond the informational stage of the CBAM model. To accomplish this movement, we need to analyze the information being communicated about CBA to ensure its accuracy. This chapter presents key discriminations that we believe implementors must make for

informational needs to be satisfied within the CBAM model. Additionally, given adoption of a scientist-practitioner model, we will identify the pieces of information and data that must be generated to validate empirically the various CBA strategies. This chapter therefore concludes with our analysis of future research needs.

Table 1

A sampling of articles on curriculum-based assessment published in refereed journals through 1989.

Journal	Authors
American Educational Research Journal	Fuchs, Deno, & Mirkin, 1984
Diagnostique	Fuchs, Deno, & Marston, 1983; Marston, Fuchs, & Deno, 1986
Exceptional Children	Blankenship, 1985; Deno, 1985; Fuchs, Fuchs, & Deno, 1985; Galagan, 1985; Gickling, & Thompson, 1985; Marston, & Magnusson, 1985; Rosenfield, & Rubinson, 1985; Tucker, 1985
Focus on Exceptional Children	Deno, & Fuchs, 1987
Journal of Behavioral Assessment	Good & Shinn, in press; Mirkin, Deno, Tindal, & Kuehnle, 1982
Journal of Educational Research	Fuchs, Fuchs, & Tindal, 1986b; Tindal, et al., 1985
Journal of Learning Disabilities	Shinn, Ysseldyke, Deno, & Tindal, 1986
The Journal of Special Education	Fuchs, & Fuchs, 1986b; Marston, 1988
Journal of Special Education Technology	Fuchs, Deno, & Mirkin, 1983
Learning Disability Quarterly	Deno, Wesson, & King, 1984b; Shinn, Tindal, Spira, & Marston, 1987; Wesson, King, & Deno, 1984
Professional School Psychology	Shinn, Tindal, & Stein, 1988
Reading Research Quarterly	Fuchs, Fuchs, & Deno, 1982
Remedial and Special Education	Fuchs, 1986; Fuchs, & Fuchs, 1984; Shinn, & Marston, 1985; Tindal, Shinn, & Germann, 1987
School Psychology Review	Deno, 1986; Howell, 1986; Neisworth, & Bagnato, 1986; Shinn, 1986; Shinn, 1988; Shinn, Rosenfield, & Knutson, 1989
TEACHING Exceptional Children	Deno, Mirkin, & Wesson, 1984a; Wesson, 1987

ANALYSIS OF THE PUBLISHED LITERATURE

Analysis of the CBA Published Literature by Its Creators

As of January 1, 1990, over 100 articles, book chapters, or books have been published investigating or describing the use of CBA strategies (for a partial listing, contact the authors). The publication channels have included, but are not limited to, all the major special education journals and most school psychology journals. With the exception of a limited set of journals such as the *American Educational Research Journal* and the *Journal of Behavioral Assessment*, few articles about CBA have been published outside of these professional domains. A sampling of journals and prominent CBA authors is presented in Table 1.

Journal articles are supplemented by an increasing number of books, including ones by Hargis (1987); Idol, Nevin, and Paolucci-Whitcomb (1986); Howell (Howell & Kaplan, 1980; Howell & Morehead, 1987); Bagnato, Neisworth , & Munson (1989); Salvia and Hughes (1989); and Shinn (1989a), as well as training monographs/materials by Gickling and Havertape (1981) disseminated by the National Association of School Psychologists.

Analysis of the CBA Published Critiques

The authors listed in Table 1 account for more than 95% of the research and scholarly articles written about CBA. An exhaustive review process failed to identify many articles written about CBA by persons other than these, although a number of resources (e.g., Will, 1986; 1989) mentioned CBA as a positive strategy. Among the eight articles that provided more than a cursory recommendation about the use of CBA, five were published in refereed journals (Reschly, 1988; Reschly et al., 1988a; Reschly et al., 1988b; Reschly et al., 1988c; Taylor, Willits, & Richards, 1988), one was a book chapter (Lentz, 1988), one was an article published in the newsletter of the National Association of School Psychologists (Lombard, 1988a) that was based on a paper presented at a state conference (Lombard, 1988b), and one was a letter to the editor in the NASP newsletter (Coates, 1989).

In an article describing the future of school psychology, Reschly (1988) proclaimed CBA as one of the most important new competencies required for school psychologists in alternative service delivery systems. He described CBA as educational assessment tools derived from a behavioral assessment paradigm where behavior is measured directly in the natural (i.e., classroom) environment. CBA was presented as a precise methodology for "measuring target behavior, monitoring

progress, and assessing outcomes" (p. 471). Further, Reschly suggested that CBA facilitates instruction on relevant skills. His description of CBA concluded with two caveats. First, professionals need specific training on CBA, as it is not a simple methodology. Second, to avoid misconceptions, it must be remembered that CBA is *not* (our emphasis) an intervention. Reschly, Kicklighter, and McKee (1988a; 1988b; 1988c) also commented favorably on CBA in a series of articles summarizing federal court cases on assessment and disproportionate placements in special education. In reviewing the rulings from the Marshall et al. vs. Georgia case (1984), they concluded that "the kind of assessment fostered by the Marshall Court is what has been called curriculum-based assessment.... CBA and other direct measures of functioning are preferable because the (assessment) results are related to interventions beneficial to the individual" (p. 20).

A more extensive critique of CBA was provided by Taylor, Willits, and Richards (1988) in an article published in *Diagnostique*. In describing CBA, Taylor et al. proposed that it was not really a new concept, and in fact, simply "formalized a long standing practice" (p. 15). CBA was essentially criterion-referenced testing (CRT) where curricular objectives were operationalized into tests and cutting scores were used to determine mastery. Many of Taylor et al.'s criticisms therefore centered on the weaknesses of CRTs. Foremost among the criticisms was that of the limited utility of CBA in assessment and decision-making practices. As stated by Taylor et al., "It is clear that CRTs alone are not sufficient to serve the many and diverse purposes of assessment. Consequently, it is doubtful that CBA will either" (p. 15). As a result of their purported limited utility, Taylor et al. recommended that CBA should be used only as a *supplemental* assessment strategy and should not supplant traditional assessment methods.

Taylor et al. went on to detail a number of other concerns about CBA. Among them, concern was expressed that the use of CBA for writing Individualized Education Plan (IEP) objectives would be a "loss of the individual" and that the content of the CBA test would dictate the content of instruction. Taylor et al. also noted concerns that the assessment procedures derived from a curriculum could not be valid if the curriculum was not valid. We assume that *valid* in the last use was used as a synonym for *effective*. Relatedly, concern was expressed that a curriculum (and thus, CBA) may not reflect the needs of special education students. Other criticisms centered on CBA's use of local norms and the technical adequacy (i.e., reliability, validity) of the measures themselves. With respect to the former, Taylor et al. argued

that the local norms developed for CBA would be difficult to interpret and would result in special education students' change of eligibility, depending on the school system in which they were enrolled. Taylor et al. (1988) concluded their critique of CBA with its positive use only in the following set of conditions:

1. If the curriculum on which the CBA is based is valid.
2. If the curriculum on which the CBA is based represents the needs of the special education student.
3. If the CBA instrument can be developed to yield reliable and valid results.
4. If limitations are acknowledged or additional research is conducted regarding the curricular areas for which CBA is appropriate.
5. If limitations are acknowledged regarding the use of CBA as a comprehensive assessment approach.
6. If careful attention is given to properly training users of CBA.

In his chapter on direct observation and measurement of academic behavior, Lentz (1988) describes CBA as employing direct measures of academic behavior that are essential to the resolution of academic problems in the classroom. CBA is seen as oriented to the determination of special education eligibility, setting individual educational plan (IEP) goals, and monitoring progress using procedures that were designed to offset the problems with "norm-based achievement tests" (p. 84). Tests are short-duration probes that assess the academic skills taught within the classroom using stimulus materials from the instructional curricula. In contrast to criterion-referenced tests, CBA procedures are used in a repeated fashion. While noting these strengths, Lentz provided a number of criticisms of CBA from a behavioral perspective. Among the criticisms was his contention that CBA research was conducted out of a nonbehavioral, psychometric approach where probes are high-inference measures about global constructs. Lentz also took issue with the use of CBA probes for problem identification/ screening as a process that "does not fit a behavioral model very well" (p. 103). Finally, he criticized CBA for its lack of utility in specifying which treatments will work. As stated by Lentz (1988), "It seems clear that CBA probe data cannot be used unilaterally to predict success of interventions" (p. 106).

The most critical review of CBA was written by Lombard (1988a). In critiquing one type of CBA, curriculum-based measurement (CBM), he asserted that it had not lived up to its promise as a "new and improved paradigm to meet special education students needs" (p. 20).

Lombard's major criticisms fell into two major categories: (a) the components of what was measured and (b) the use of the measures for purposes in making special education eligibility decisions. His concerns about what comprised the CBM probes were similar to those cited by Taylor et al. (1988), including curriculum bias, speed effects, effects of students' attentional and psychomotor deficits on their scores, and what he referred to as the tests' limited behavior sampling. Lombard's concerns about CBA were directly counter to the Reschly et al. interpretation of the *Marshall* (1984) court case. Lombard expressed concern that CBA strategies were both discriminatory towards minorities and would redefine the special education population by placing low-achieving, not-truly-handicapped students in special education. Further, he stated that the use of CBA has allowed the general education system to "short-cut" the requirements of PL 94-142.

The final critique by non-CBA authors was that of Coates (1989). In his brief but succinct commentary, Coates praised curriculum-based assessment as an exciting new measurement technology. However, he also raised concerns about the apparent assumption of many CBA proponents that standardized norm-referenced tests have no usefulness beyond placement decisions and the notion that norm-referenced testing and CBA are antagonistic, as well as concerns about the validity of CBA reading measures.

CBA Informational Needs for Educators

How does one reconcile the differences in interpretations and criticisms of CBA by authors such as Reschly, Taylor, Lentz, Lombard, and Coates? If Hall and Hord's concerns-based adoption model is employed, what current informational needs are suggested to allay personal concerns and facilitate implementation of this innovative technology? Based on our analysis and knowledge of the published CBA references and the criticisms of CBA, we see the need to engage in a series of *discriminations* within the existent literature, including distinguishing between (a) assessment terms, (b) assessment decisions, (c) different models of CBA, (d) assessment paradigms, and (e) CBA-based changes and the change process itself.

Discriminating Between Assessment Terms

The easiest discrimination that can be made within the existent literature on CBA is to clarify the terms that are used to describe both CBA and other measurement tools. We have observed the terms *assessment, standardized, norm referenced, criterion referenced, informal,*

formal, and *published* to be bandied about almost casually, and often interchangeably. We propose that all authors increase the precision of the language used to describe various measurement terms. As two cases in point, consider the term *norm referenced* as used by Coates (1989) and Lentz (1988). Coates asserted that CBA is, in a sense, against "standardized norm-referenced" tests. Lentz described CBA as a system developed to overcome problems with "norm-based" achievement tests. In both cases, the authors are referring to commercially available, norm-referenced achievement tests. The key term is *commercially available*, not *standardized* or *norm referenced*. CBA can be standardized (i.e., administered and scored in a prescribed, replicable manner) and can be used in a norm-referenced manner where a specific student's score is compared to a normative sample (Shinn, 1989b). The use of terms *informal* and *formal*, with the former implying either nonstandardized and/or not commercially available and the latter implying standardized and/or commercially available and/or norm referenced, contribute little information and less ambiguous terms are available. We believe the salient features of academic assessment can be described using the following terms and definitions:

1. *Standardized:* A *test* that is administered and scored in a specified, replicable manner.
2. *Nonstandardized:* A *procedure* for collecting data that is idiosyncratic to the examiner, with results that may have little generality across individuals and time.
3. *Commercially available:* A test *or* procedure that is produced by a publisher.
4. *Norm referenced:* A *test* that has interpretive metric(s) derived from a comparison group.
5. *Criterion referenced:* A *test* that has items derived from an identified instructional domain, with interpretive metric(s) derived rationally (i.e., without sampling from a group of students).
6. *Individually referenced:* A *test* that has items derived from an identified, finite instructional domain, with interpretive metric(s) derived by comparing the student's score to his or her previous scores over time.

All *tests* are standardized. Single terms thus may be used hierarchically. For example, a published, norm-referenced test (Woodcock Reading Mastery Test) implies, by definition, standardization. These distinctions can eliminate many confusions engendered by authors.

Discriminating Between Models

The articles by Reschly (1988) and Taylor et al. (1988) provide clear evidence of the need to clarify that CBA is not a unified set of procedures or strategies. There is no *one* model of CBA. Although generally quite accurate in his description of CBA, Reschly (1988) errs in stating categorically that CBA is behavioral assessment applied to academic problems and that CBA is not an intervention. The specific accuracy of his statements is dependent upon which model of CBA is considered. Models of CBA range from those placing great reliance on a behavioral assessment paradigm (Deno, Mirkin, & Shinn, 1979; Knutson & Shinn, 1990; Shinn, Goodwin, & Habedank, 1989) to those that are decidedly nonbehavioral (Gickling & Havertape, 1981). With respect to the contention that CBA is not an intervention, it is important to note that *all* assessment, including CBA, is to some degree an intervention; data are derived to improve the functioning of the individual assessed. The degree to which CBA is or is not an intervention parallels the continuum of whether the CBA model is behavioral. The model of CBA represented by Deno currently represents the end of the continuum where it is less of an intervention. Gickling's model, on the other hand, represents the other end of the continuum, as it is almost exclusively an intervention strategy.

Errors of discrimination between models are made also by Taylor et al. (1988). As presented earlier, these authors consider CBA to be essentially criterion-referenced testing (CRT) where a curricular objective is identified and a test and mastery score are constructed to correspond to the domain that the objective represents. CBA is treated as synonymous with CRTs and Taylor et al. view it as having the same strengths and weaknesses. However, it is apparent from an examination of the reference list for the Taylor article that the authors are referring to five *different* models of CBA. We have classified the types of CBA model and authors in Table 2. Only *two* of these models, the Blankenship CBA-CRT and the Bagnato, Neisworth , and Munson preschool CBA model, could be characterized as CRTs. The other models are not based on traditional conceptions or definitions of criterion-referenced testing. Although all derive their testing items from the curriculum, the accuracy-based model of CBA, Curriculum-Based Measurement, and CBA for instructional design do not create CRTs for each curricular objective, nor do they establish mastery criteria on a rational basis.

Table 2
Classification of the different CBA citations characterized as one CBA model
in Taylor, Willits, and Richards (1988) into different models of CBA.

Author(s)	CBA Model
Blankenship, 1985	Criterion-referenced CBA
Bursick, & Lessen, 1987	CBA for Instructional Design
Deno, 1985; Fuchs, & Fuchs, 1986b; Lombard, 1988a; Lombard, 1988b; Marston, & Magnusson, 1985; Shinn, 1988; Wesson, King, & Deno, 1984	Curriculum-Based Measurement
Coulter, 1985; Rosenfield, & Rubinson, 1985	Accuracy-Based CBA
Neisworth, & Bagnato, 1986	CBA for preschool assessment

A growing number of professional resources are available that provide information for professionals to discriminate between the differing models of CBA (Marston, 1989; Shinn, Rosenfield, & Knutson, 1989; Tindal, this volume). It is beyond the scope of this chapter to detail sufficiently the important differences among CBA models. Suffice it to say that it is critical to discriminate among models. Failure to do so increases the likelihood of misunderstandings by practitioners. According to Hall and Hord (1984), lack of good information will impair resolution of the self-concerns in the systems-change process. It is important to note that discriminating among models does not imply incompatibility. Shinn, Rosenfield, and Knutson (1989) have argued that although the CBA models differ in some important ways, they have the potential to fit together to form a coherent problem-solving educational assessment system. Without discriminating between models, however, practitioners run the risk of overgeneralizing. In particular, they may misinterpret criticisms of *one* specific CBA model as pertaining to *all* CBA procedures. Technical adequacy (i.e., reliability, validity) is a case in point. Taylor et al. (1988) raised concerns about the technical adequacy of CBA. A novice in CBA may interpret Taylor's statement to be applicable to all models of CBA when one model, Curriculum-Based Measurement (CBM), has extensive documentation of its technical adequacy.

By combining injudiciously those features of the various CBA models that are genuine weaknesses, that are undeveloped (e.g., secondary applications of CBM), or that are beyond the intended focus of the model (e.g., school-age applications of Bagnato, Neisworth, and Munson's Preschool CBA), critics and practitioners can create the educational equivalent of an Edsel: a measurement and decision-making system that is indefensible. Alternatively, we believe firmly that selecting and combining specific strengths from across CBA models in practice can generate the educational equivalent of a Mercedes-Benz.

Discriminating Between Assessment Decisions

In general, most assessment practices suffer from a lack of distinguishing what decision is to be made with the data. Although the use/overuse of published, norm-referenced tests (PNTs) is most frequently the target of criticism in this regard (Salvia & Ysseldyke, 1987), CBA also suffers for similar reasons (Shinn, Rosenfield, & Knutson, 1989). There appears to be a high likelihood of overstating the utility of the data derived from any test. As a result, we witness the continued practice of trying to plan instructional programs from PNTs, despite a lack of data to suggest that they can be used for such purposes (Deno, 1986). Similarly, we see some models of CBA being described as a "do-it-all" approach *without data* to do so. In order to select the most appropriate assessment procedure, one must first ask, "What decision am I being asked to make?" The demands placed on an assessment device vary with the educational decision being made.

Regardless of the strategies used to derive student data, we believe that assessment practices will be improved only when viewed within a decision-making context. Salvia and Ysseldyke (1987) have provided one decision-making model where data are collected to facilitate screening, eligibility determination, intervention planning, pupil progress, and program evaluation decisions. Their heuristic provides a mechanism by which assessors can select strategies for collecting data to make decisions. In recent years, we have adopted a decision-making paradigm that closely approximates that of Salvia and Ysseldyke. Within a *problem-solving paradigm*, educational decisions are classified as problem identification, problem certification, exploring alternative solutions, evaluating solutions, and problem solution. The first four of the decisions correspond roughly to those of Salvia and Ysseldyke. When the last decision, problem solution, is added, one has a framework for making decisions about individual students that is less student centered and more situation centered than the Salvia and Ysseldyke

paradigm (for a more detailed discussion, see Shinn, Nolet, & Knutson, 1990). Within a problem-solving model, a problem is defined as a difference between what is expected and what occurs. Each step of the problem-solving model specifies a measurement strategy (the data to be collected) and an evaluation strategy (the decision to be made). The measurement and evaluation activities, as well as specific data collection strategies within the problem-solving model, are summarized in Table 3.

Table 3
Summary of Problem-Solving Model Decisions, Measurement Activities, and Evaluation Activities.

Problem-Solving Decision	Measurement Activities	Evaluation Activities	Specific Tasks
Problem Identification	Record Differences Between Expectations and Student Performance	Does a Discrepancy Exist?	Peer-Referenced Assessment
Problem Certification	Describe Severity of Discrepancy and Available Resources in Environment That Many Reduce Discrepancy	Are Additional Services Beyond Those Currently Available in the Typical Environment Needed?	Survey-Level Assessment & Evaluation of General Education Modifications
Exploring Solutions	Estimate Expected Student Gains and Available Alternative Resources	Which Intervention Will Be Implemented? What Are The Intervention's Goals	Write Long-Term Goals, Design Intervention Plan
Evaluating Solutions	Monitor Program Intervention, Student Progress	Is Program Effective, Is Student Making Progress?	Collect Data, Compare Actual & Expected Performance
Problem Solution	Record Differences Between Expectations and Student Performance	Are Additional Resources Still Needed To Reduce Discrepancy	Repeat Peer-Referenced Assessment

Adapted from S. Deno (1989). Curriculum-Based Measurement and Special Education Services: A Fundamental and Direct Relationship. In M.R. Shinn, (Ed.) Curriculum-Based Measurement: Assessing Special Children, (pp. 1-17). New York, NY: Guilford Press.

Problem Identification and Certification place a high reliance on norm-referenced data to operationalize the severity of the discrepancy between what occurs and expectations. However, norm-based assessment strategies are less than useful for Exploring and Evaluating Solutions. Failure to discriminate between the decisions to be made and the data to be collected can result in inappropriate and ineffective assessment practices. Given the considerable differences that exist between CBA models with respect to their evidence for decision-making utility, failure to make these discriminations is likely to be common and problematic.

Discriminating Between Assessment Paradigms: Current and Problem-Solving Educational Assessment Practices

A key discrimination that must be made in this discussion is between CBA as an assessment technique (i.e., CBA as another "test") and the *paradigm* used to select and evaluate assessment techniques. The problem is not just that CBA techniques provide different data to answer the questions schools ask. Instead, we suggest that CBA *may* address different questions based on different underlying assumptions and values; in other words, a different paradigm. We add the caveat *may* in that, with the exception of CBM, the assumptions and values underlying most models of CBA have yet to be made explicit. The assumptions, philosophical underpinnings, and values specified overtly for CBM (e.g., Deno, 1985; 1986; 1989) clearly demonstrate fidelity to a different educational assessment paradigm, of which CBM is an important, but not the sole, component (Deno, 1989; Knutson & Shinn, in press). Our discussion of paradigm shift will focus, therefore, on the CBM model of CBA and the problem-solving paradigm.

We suggest that discussions of the value and future of CBA occur at two levels of discourse: paradigm and procedure. At the paradigm level are the values, assumptions, and regularities of current practice that generate the criteria by which we evaluate the adequacy of assessment techniques. At the procedure level is the evaluation of specific techniques or procedures with respect to established criteria. At the procedure level, we might ask, "How good is this assessment technique?" At the paradigm level, we might ask, "How will we know a good technique when we see one?" The paradigm/procedure distinction is crucial because decisions about quality are based on different types of information at each level. Technique questions are resolved empirically by comparing the extent to which alternative

procedures satisfy established assessment criteria (e.g., best reliability, strongest criterion-related validity). In contrast, paradigm conflicts are resolved on the basis of values and assumptions. What purpose should we be trying to accomplish with our assessments? *Why* do we want to accomplish this purpose? Data are involved only in more general terms, as broad strokes of the research brush regarding the empirical support for underlying assumptions.

The distinction between procedure and paradigm is important because educators are questioning both levels. With respect to the former, attention is focused on the technical adequacy of current CBA assessment techniques. With respect to the latter, professionals are struggling with the larger issue of what is the "best" or "right" way to make data-based decisions about students. We argue that the future of CBA is not dependent solely upon procedure but is entwined inextricably with resolving what is the best way to make assessment decisions. If CBM is used merely to accomplish the same goals and objectives as current techniques, based on the same underlying values and assumptions (i.e., as a *supplement* to current assessment techniques) with more content-valid devices, its future most likely will be short, and perhaps deservedly so. Practitioners already are experiencing difficulty keeping up with their caseloads and, most likely, additional time and assessment requirements will not be received with enthusiasm. Further, it is likely that assessment activities will continue to be used only for child-find, special-education-eligibility decisions and not to improve student outcomes.

Paradigm questions must be resolved before assessment procedures can be compared meaningfully. In order to evaluate the worth of an assessment technique, we must first determine the purposes we expect the procedure to accomplish and clarify the rationale for those purposes. Only when the goals and purposes of assessment are established can we compare how well alternative assessment procedures accomplish those goals. Comparing current and alternative paradigms requires clarification of the values, assumptions, purposes, and goals of assessment. Unfortunately, the current assessment paradigm is not well articulated, so discussions of paradigm shift are difficult.

To illustrate the implications of a paradigm shift, we have constructed our best understanding of the current assessment paradigm based on the *existing regularities* found in current practice. An examination of existing regularities is important from a systems-change perspective. Sarason (1982) asserts that for change in schools to take place, one must make two assumptions: (a) that the change is desirable according to some set of values and (b) that the intended outcomes are

clear. Sarason (1982) maintains that the implied outcomes of any change process are "changing the existing regularity, eliminating one or more of them, or producing new ones" (p. 96). A regularity is a programmatic or behavior occurrence that is supposed to have an intended outcome. It is often an unspoken, assumed belief that is not data based. One regularity cited by Sarason as an example is that generally children in this country go to school 5 days per week (Monday through Friday). Often, however, the intended outcome of the regularity itself (as in the previous example) may not be clear, and there frequently are no systems built into schools to ascertain the discrepancy between regularities and intended outcomes (Sarason, 1982).

Important existing regularities implicit in current assessment practices are compiled in Table 4. We do not assume this list of regularities to be exhaustive. These regularities impact both the information we attempt to obtain and the criteria by which we evaluate the quality of assessment techniques. Within the regularities are implied anticipated outcomes, social values, and methodological testing techniques.

Table 4
Regularities Questioned by Implementation of Curriculum-Based
Assessment as Embedded Within Problem Solving.

1. Commercially available, norm-referenced tests are used mostly by psychologists in a diagnostic-perscriptive fashion to identify, in advance of treatment, the interventions that will be successful (Deno, 1986).

2. Commercially available, norm-referenced tests are used in a pre and post testing format, usually on a yearly basis by teachers to evaluate student progress and intervention effectiveness (Deno, 1986).

3. Group designs are used for making statements about the effects of individual student programs (Deno, 1986).

4. Instruction not individualized nor evaluated. An assumption is made that what works for one student works for all students (Deno, 1986).

5. Students only are examined intensively because they are the cause of academic problems (Alessi, 1989).

6. Handicapping conditions (e.g., learning disabilities) are identified by school psychologists' testing students using commercially available, norm-referenced tests (Heller, Holtzman, & Messick, 1982).

7. We don't evaluate alternative interventions (e.g., special education) systematically because we know they are effective and therefore do not need to be evaluated (Deno, 1986).

If existing regularities are to be changed, the outcomes, values, and assumptions must be examined explicitly to determine whether there is a defensible underlying paradigm and whether an alternative paradigm should be adopted. We have attempted to translate the existing regularities into 10 dimensions of assessment practices that embody a paradigm. These dimensions are presented in the first column of Table 5. In column 2, questions that allow one to determine the quality of the practice are provided for current assessment procedures. The evaluative questions in column 2 are drawn from classical test theory and standard instruction in tests and measurement. With regard to the *purpose* of assessment, for example, if the existing regularities are to group students by handicapping condition and to provide corresponding interventions (e.g., special education services) on the basis of published, norm-referenced tests, assessment techniques must discriminate among students reliably. Assessment techniques that generate spread or variability in individual performance consequently are judged more apropos than those that do not. The intended outcome presumably is to provide appropriate instruction and services to children grouped by their classification. That this is an assumption or belief and not a data-based outcome is evidenced by the pervasive difficulties documenting the efficacy of special education placement (Heller, Holtzman, & Messick, 1982), and the regularity that interventions are not evaluated systematically.

Earlier, we reported Sarason's (1982) contention that for school change to occur, it must be desirable based on some values. We believe that the professional values espoused by school psychology leaders (e.g., Bardon, 1988; Graden, Zins, & Curtis, 1988; Reschly, 1988), as well as the results of the most recent survey of NASP leaders and practitioners (Reschly, Genshaft, & Binder, 1987), suggest that change in the current assessment paradigm is desired. However, we also believe the outcomes of alternative assessment practices have not been examined with regard to the changes that would be required in existing regularities. Although widespread dissatisfaction has been expressed with the current assessment paradigm, there is as yet no consensus regarding the preferred alternative assessment paradigm.

The alternative assessment system we propose is problem-solving educational assessment. In this paradigm, the ecological educational assessment model described by Shapiro and Lentz (1985) and the behavioral assessment model described by Barlow, Hayes, and Nelson (1984) are integrated within the problem-solving sequence detailed by Deno (1989) presented earlier. The model also addresses advances in and extensions of classical test theory (e.g., Messick, 1989). Knutson

Table 5
Different Questions Resulting from Paradigm Shift.

Dimension	Current Assessment Paradigm	Problem-Solving Paradigm
Purpose	Do assessment results spread out individuals facilitating classification/placement into groups?	Does assessment result in socially meaningful student outcomes for the individual?
Test Validity	Does the assessment device measure what it says it measures? Criterion-Related Validity: Does the test correlate with other tests purporting to measure the same thing? Construct Validity: Does the test display a stable factor structure?	Are the inferences and actions based on test scores adequate and appropriate (Messick, 1989)? Treatment Validity: Do decisions regarding target behaviors and treatments based on knowledge obtained from the assessment procedure result in better student outcomes than decisions based on alternative procedures (Hayes, Nelson, & Jarrett, 1986)?
Unit of Analysis	Groups: Probabilistic statements about individuals: Do students with similar assessment results most likely display similar characteristics?	Individuals: Does assessment show that this treatment is working for this student?
Time Line	Summative: Does the assessment indicate whether or not the intervention Did work?	Formative: Does the assessment indicate whether or not the intervention Is working?
Level of Inference	Does the assessment provide an indirect measure of an unobservable construct?	Does the assessment directly measure important target behaviors or skills?
Locus of the Problem	Does the assessment identify relevant student characteristics that contribute to problem etiology?	Does assessment identify relevant curriculum, instruction, and contextual factors contribute to problem solution?
Focus	Problem Certification: Does assessment accurately identify problems?	Problem Solution: Does the assessment accurately identify solutions?
Test Reliability	Are test scores stable over time? Are scores based on different behavior samples, obtained in different contexts/settings consistent?	What factors account for the variability in student performance?
Context	Does the assessment provide a comparison with students receiving a nationally representative range of curriculum and instruction?	Does the assessment provide a comparison with students receiving comparable curriculum and instruction?
Dimension of Dependent Variable	Does the assessment provide information regarding the level of pupil performance?	Does the assessment provide information regarding the level of pupil performance and the slope of pupil progress?

and Shinn (in press) provide details as to how the problem-solving educational assessment paradigm is operationalized. The evaluative questions within a problem-solving model by dimension are presented in column 3 of Table 5.

We believe that for a paradigm shift to occur, we must contrast current and alternative assessment practices by their evaluative criteria within each dimension. The juxtaposition of assessment questions in Table 5 illustrates the fundamental and far-reaching differences in assessment resulting from a paradigm shift. To illustrate in more detail some of the fundamental differences between paradigms, we will contrast the use of intelligence tests in decision making with instructional problems and CBM within a problem-solving model. In current practice, intelligence tests are used frequently to assist in decision making about academic problems. A major purpose purportedly is to provide a prediction of future learning. Educators might want to evaluate, for example, "a student's ability to benefit from instruction." If inadequate learning or academic progress is predicted to occur as a result of the student's ability to benefit from the types of instruction available within general education settings (e.g., the student obtains an IQ below 70), the student customarily is identified as handicapped and special education services are recommended. With the instruction available in special education settings (i.e., individualized educational programs, modifications in the curriculum and instruction), the student is anticipated to make better academic progress.

Within the problem-solving paradigm outlined in Table 5, practice would differ substantially. A problem would be defined as a discrepancy between observed and expected behavior (Deno, 1989). Assessment would examine the student's academic progress in curricular material over time. If the level of student skills or the rate of student progress was not adequate, alternative interventions would be implemented and evaluated systematically. Interventions would include modifications of instruction, curriculum, and context variables not necessarily requiring special education services. Interventions resulting in improved academic progress would be maintained and modified. Perhaps more importantly, interventions that were ineffective for the individual student would be changed. From this perspective, the assessment of intellectual functioning does not contribute to educational decision making.

Using the assessment of intellectual functioning as an exemplar, the effects of a shift in paradigms are examined with respect to the dimensions of the dependent measure, the level of inference, the unit of analysis, and the context of assessment.

DIMENSION OF THE DEPENDENT VARIABLE

A fundamental difference between assessment paradigms regards the dimension of the dependent variable. The current assessment paradigm features a one-dimensional view, stressing a static measure of the *level* of pupil skills only. The problem-solving paradigm includes a second dimension of performance—time—stressing a dynamic examination both of the level of pupil performance and the *slope* of pupil progress.

Considerable confusion exists in the professional literature between the assessment of slope and level. The level of pupil performance refers to the amount or extent of skills displayed by the student at one point in time. An estimate of level is obtained from one assessment. The slope of pupil progress refers to the rate at which the student is acquiring skills over time. Obtaining an estimate of slope requires repeated assessments of skill level over time and a procedure for summarizing the rate of change (Good & Shinn, 1990; Shinn, Good, & Stein, 1989). From a mathematical perspective, slope refers to the unit change in a dependent variable (Y) associated with a unit change in an independent variable (X):

$$\text{Slope} = \frac{Y_2 - Y_1}{X_2 - X_1} \tag{1}$$

Because intelligence tests typically are given in one sitting at one point in time, IQ tests are, by definition, measures of the level of pupil performance only. On this day, Billy obtained an IQ score of 85 on the WISC-R. This outcome means that on this day, on these tasks, and under these conditions, Billy displayed skills at a level of proficiency one standard deviation below the mean. In contrast, a problem-solving paradigm would stress the assessment of skills over time. Using CBM, for example, a student's skills would be assessed on a frequent, repeated basis, with the results plotted on a two-dimensional graph (time by level of skill). The slope of pupil progress then would be used to evaluate the efficacy of interventions and the need for alternative, potentially more intrusive, interventions.

LEVEL OF INFERENCE

A second fundamental difference between assessment paradigms regards the level of inference entailed in decisions about individual students. In general, when compared to the problem-solving paradigm, the current assessment paradigm countenances a much higher level of inference as decisions are based on less observable constructs and less direct data, and entail more assumptions that are more difficult to substantiate or are less tenable (Kratochwill & Shapiro, 1988). As discussed previously, intelligence tests are measures of students' level of performance. However, they typically are used to make high inference statements about the future slope of pupil progress. When educators use an IQ test to determine a "student's ability to benefit from instruction," for example, they are making an inference about the slope of pupil progress. Substantial benefit corresponds to a steep slope; little benefit corresponds to a shallow slope. Indeed, many researchers define intelligence (i.e., ability or aptitude) in terms of slope. Carroll (1989), for example, notes that "aptitude is the name given to the variable or variables that determine the amount of time a student needs to learn a given task, unit of instruction or curriculum to an acceptable criterion of mastery under optimal conditions of instruction and student motivation" (p. 26). Thus, under fixed conditions of instruction, the student with higher ability would display the steeper slope of pupil progress (i.e., acquire skills in a shorter length of time). The correspondence of IQ to slope of pupil progress also is evident in the familiar formulation of the ratio IQ, the initial metric of intelligence tests. The ratio IQ is defined as:

$$\frac{\text{Ratio IQ}}{100} \qquad \frac{\text{MA}}{\text{CA}} \qquad \frac{\text{MA} - 0}{\text{CA} - 0} \qquad (2)$$

Or, alternatively, as:

$$\frac{\text{Ratio IQ}}{100} \qquad \frac{Y_2 - Y_1}{X_2 - X_1} \qquad (3)$$

Thus, the ratio IQ represents the amount of change in intellectual skills associated with a unit change in time over the individual's entire life

span, or the slope of pupil progress on intellectual, problem-solving skills. Clearly, then, statements about the slope of pupil progress are one intended purpose of intellectual assessment.

The use of intelligence tests to make inferences about future learning is not altogether unreasonable. However, meaningful conclusions about the slope of pupil progress may be drawn from measures of the level of pupil performance (e.g., an intelligence test) only when appropriate assumptions are met. As illustrated in Figure 1, inferences about slope based on comparisons of level require four assumptions. First, students must be at the same level at the beginning of the relevant time period (Time$_1$). For the ratio IQ, the implied time period begins at birth (CA = 0) where, indeed, intellectual skills conceptually are identically 0. When shorter time periods are considered, as in the student's educational career or the current academic year, the assumption of equal entry levels is more difficult to support. If students display different entry-level skills, different final-level skills would *not* be indicative of differences in slope.
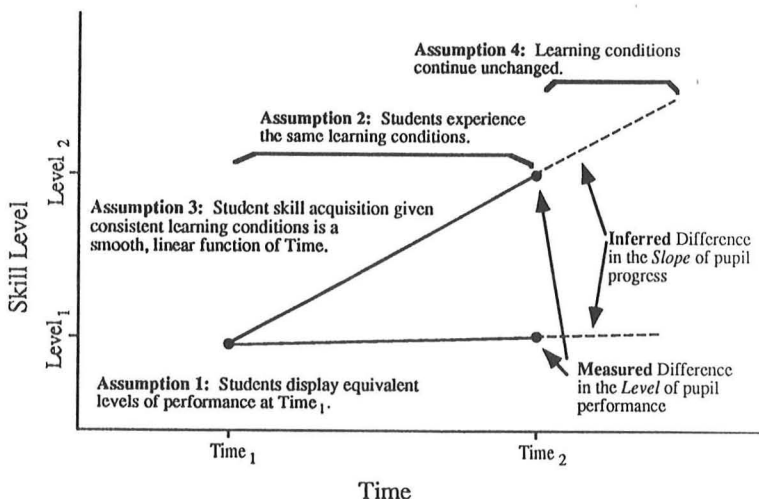
Figure 1. A graphic representation of the assumptions required
to use a measure of level to infer slope.

The second assumption is that the students experienced identical learning conditions. To the extent that instructional conditions impact the slope of pupil progress (i.e., learning), different conditions would be confounded with differences in slope. Under disparate learning conditions, differences in the level of pupil skills could represent differences in the quality of instruction rather than a child characteristic.

The third assumption is that the acquisition of skills is a smooth, linear function of time, given consistent instruction. To the extent that the slope of pupil progress is sporadic or nonlinear, previous slope, especially over long time periods, would be less related to current or future slope. The fourth assumption is that learning conditions continue unchanged. A change in learning conditions would be expected to impact the slope of pupil progress, rendering inferences about current and future slope invalid.

Only when all four assumptions are tenable can inferences about the slope of pupil progress be made from differences in the level of pupil performance. When inferences about the slope of pupil progress in an academic content area are based on differences in the level of intelligence test performance, an additional, fifth assumption is necessary. This additional assumption is that the slope of pupil progress is consistent across skill areas. In particular, the slope of pupil progress on the tasks sampled by the intelligence test is assumed to be the same as the slope of pupil progress on academic skill measures, like oral reading fluency.

Clearly, making decisions about the slope of pupil progress based on intelligence test performance is a high-inference activity, requiring multiple assumptions that are difficult to assess and that vary in plausibility. It is no surprise that the few studies examining empirically the relationship between the slope of pupil progress and level of intellectual functioning have found little or no relationship (Bailey, 1981).

In contrast, a problem-solving educational assessment paradigm emphasizes a substantially lower level of inference. By assessing pupil progress *directly* in the skill area of interest, it is not necessary to assume that the slope of pupil progress is consistent across skill areas. By basing educational decisions on repeated measurements of academic skills over time, slope can be observed instead of inferred. It is not necessary to make extensive assumptions about instructional conditions and beginning skill levels. In addition, the conclusions drawn are at a much lower level of inference: At this time, under these instructional conditions, the slope of pupil progress was not adequate. Slope of pupil progress is not considered a student characteristic only, but is instead a combination of the student and the conditions of instruction. This approach requires a low-level assumption that the slope of pupil progress will continue unchanged in the absence of a change in instruction, curriculum, or conditions.   However, a change in instructional conditions is not assumed to increase the slope of pupil progress. Instead, the slope of pupil progress following an intervention again is assessed.

UNIT OF ANALYSIS

A third, fundamental difference between assessment paradigms regards the unit of analysis and interpretation. The assumptions required to make inferences about the slope of pupil progress based on measures of the level of intellectual functioning may be reasonable—for groups of students. In general, students are exposed to reasonably stable, homogeneous learning conditions (i.e., school) and enter school with roughly equivalent skills. Similarly, criterion-related validity studies repeatedly have demonstrated the relationship between intelligence test performance and academic achievement, again for groups of students. As a result, one can be completely confident that a *group* of students with low intelligence test scores will experience more difficulty in school than a group of students with high scores. Individual students with low scores, however, may *or may not* experience academic difficulty. Statements about individuals based on intelligence test scores are possible on a probabilistic basis only. With the relationship between academic achievement and intellectual functioning ranging between .60 and .80, students with low intelligence test scores will display substantial variability in academic performance. Some individuals will display quite high academic skills. Macmann, Barnett, Lombard, Belton-Kocher, and Sharpe (1989) provide an excellent illustration of this problem. They show that when two measures are correlated .80, and individuals are selected on the basis of extreme scores on one measure (i.e., 1.96 standard deviations below the mean), many cases will fall at or near the mean of the second measure.

From the perspective of the problem-solving paradigm, the question is *not* whether this individual student is a member of a group that, as a group, experiences academic difficulty. Instead, the question is whether *this individual student* is experiencing academic difficulty; the unit of analysis and interpretation is the individual.

CONTEXT

The problem-solving paradigm differs substantially from the current assessment paradigm with respect to the role of context in the interpretation of assessment results. The context differences are epitomized by Taylor et al.'s (1988) arguments about local and national norms and the quality of the curriculum. These authors questioned, "How might CBA affect students performing at a satisfactory level within a school where the average student performance was considerably below average compared to other norms (national, state, or even

district)? The chances are that those students would not be identified for services even though they might need help" (p. 16). They also expressed concern that the school may not be using a "valid curriculum" (presumably one that is effective), and therefore that "CBA can be no better than the curriculum selected for instruction" (p. 17). We believe this point of view exemplifies most current assessment practices with respect to context, that a problem should reside solely within the student independent of context. Environmental expectations and characteristics, in terms of how other students perform or whether the curriculum is effective or ineffective, are not relevant to the identification of the problem. This position implies two potential outcomes: (a) that a student performing at a satisfactory level within a school where the average student performance is considerably below average compared to other (e.g., national) norms should be eligible for special education services, and (b) that a student performing considerably below expectations in his or her school but above other (national) norms should not be eligible for special education services. However, a focus on within-student pathology independent of context may be inconsistent both with best practices and with current practice.

It is crucial to examine more closely the implications of emphasizing within-child pathology independent of the context of the problem. Failure to consider context may result in untenable conclusions. In the first case, are we saying that identifying within-child pathology (e.g., learning disability or mental retardation) provides an acceptable amelioration for a dysfunctional system (e.g., ineffective curriculum)? Does this mean that the system can say five "Hail Marys," 10 "Our Fathers," place 15 children in special education, and receive absolution from the sins of its curriculum? In the second case, are we saying that we should do nothing because there is no "problem"?

In current practice, context effects on decision making regarding who receives special education services have been demonstrated empirically and repeatedly. For example, Singer, Palfrey, Butler, and Walker (1989) found in a recent study of five large school districts that districts "differed in the percentage of students they identified as handicapped, the frequency with which they used various labels, the criteria used to define groups, and the functional levels of students given the labels. Consistency was greatest for those labeled hearing impaired and, to a lesser extent, physically/multiply handicapped and weakest for those labeled mentally retarded and emotionally disturbed; results for those labeled speech impaired and learning disabled fell between these two extremes" (p. 278).

We agree that a problem exists when student performance is in the average range in the context of a school system that is substantially below average compared to national norms. We disagree that the problem is within the child or that placement in special education is the solution. Placing large numbers of students in special education will not change the fact that the school is severely below average compared to national norms and may not be providing an effective curriculum. Clearly, if the school or district is severely discrepant from national norms, the system has a schooling problem.

We also disagree that when student performance is below what is typical in a system that is above average compared to national norms, a significant problem does not exist. If the child is severely discrepant from expectations within the local context, the child may have a learning problem. For example, the child may exhibit low motivation, have poor attendance, display language difficulties, be receiving inappropriate or insufficient instruction, or be inappropriately placed in the curriculum. Individual interventions possibly necessitating special education services may be indicated.

Perhaps the future will hold a divided special education funding stream. One stream would fund services for individual students based on skills discrepant from local norms or expectations. A second stream would fund services for school systems or districts. A school district might be identified as *severely teaching disabled* (STD) based on performance discrepant from national norms or expectations. Special education services might include in-service training for teachers, improved curriculum materials, hiring incentives to attract and keep quality educators, and nutritional or early intervention programs for the community, among other possibilities.

Distinguishing Between Changes in Practice as a Result of CBA and the Change Process Itself

One of our colleagues has self-titled a law about the change process (Stoner, personal communication, 1988). Stoner's Law goes something like this: When you ask someone to change, you are asking them to do more work. Asking people to do work often makes people angry. Therefore, when you ask people to change, you will make them angry. Under the best of circumstances, change will make only half the people involved angry; under the worst of circumstances, assume that change will make 95% of those involved angry. Introduction of CBA strategies in the schools is asking people to change. Whether CBA is an improvement to existing practices may be irrelevant when viewed in

the context of Stoner's Law. Attributes aside, we argue that we will need to discriminate implementation of CBA from the reactions to any change process. We can recall one particular circumstance where a school district was engaging in a general review of assessment and decision-making practices simultaneously with introduction of CBA. It was discovered by district personnel that no observations were being conducted prior to placement of students in programs for learning disabled students as required by state law. Resolution of the situation was interpreted (by teachers who had to conduct the observations) as being caused by CBA. In another district, we observed a school psychologist who was resistant to CBA centering his opposition on non-categorical placement, a school district practice that again was outside the direct effects of the implementation of CBA. Too often, changes in roles and responsibilities in general are often attributed to the innovation itself. Implementors should expect resistance to implementation and should work carefully to separate out the larger issues from those of implementing CBA.

FUTURE KNOWLEDGE AND INFORMATION

We have taken the position that the evaluation of CBA should be based on an analysis of empirical outcomes, that useful assessment strategies should be documented to "work" in some way. An extensive body of research has been accumulated on CBA strategies in approximately 10 years. However, we are concerned that most of the empirical work has centered on CBM. Other CBA models have undergone little systematic inquiry. Many additional questions exist within CBM as well. We propose that the future information needs for successful implementation be examined in three separate areas: (a) establishment of technically adequate CBA measures, (b) use of the measures in decision making with students, and (c) research on implementation.

Establishing Technically Adequate CBA Measures

Research on CBA measures must proceed in two interrelated areas. First, the pool of available measures with demonstrated technical adequacy must be increased. Second, CBA procedures must be identified for use with specific ranges of student populations (e.g., preschool, elementary, secondary).

Technical adequacy. We believe that CBA measures must meet professional standards for quality assessment devices if they are to be used for making important decisions with children. The major strategies

by which tests' quality is determined, a nomothetic, psychometric approach, or an idiographic, behavioral assessment approach, are merging in practice so that elements of both often are offered as evidence without contradiction (Barrios, 1988).

To date, only CBM researchers have undertaken extensive empirical studies of the technical aspects of their proposed instruments. CBM measures are constrained currently to the basic skills areas of reading, spelling, math, and written expression, with decreasing knowledge of technical properties in the respective order presented here. Although robust in their use with elementary-level and middle-school-level students with basic skill problems, the primary behaviors assessed with CBM, as with any assessment device, lack usefulness for *all* students. Work has proceeded with other CBM measures of reading than oral reading fluency (e.g., maze) and written expression (Tindal & Parker, 1989).

The lack of attention to reliability and validity of the other CBA models may stem from their primary use in making instructional planning or Exploring Solutions decisions. Evolving out of teacher informal testing using curricular materials, the foremost criterion for their quality was the degree to which they matched instructional content (i.e., content validity). Some researchers (e.g., Messick, 1989) have argued that content validity is not a form of validity but is a test construction issue. We believe strongly that CBA advocates *must* go beyond content validity to support their measures' quality. To the degree to which decisions other than Exploring Solutions are made, we must provide evidence that a test is accurate (reliable) and measures what it says it measures (valid). A necessary precursor to technical adequacy is explicit specification of measurement procedures.

Application of specific CBA-model strategies across age ranges. The procedures within most CBA models currently are associated with specific age- or grade-level populations. For example, the strategies represented by Neisworth and Bagnato (1986) are used with preschoolers, whereas Gickling's measurement procedures have an elementary-grade focus. It seems worthy to consider expanding the measurement strategies associated with the philosophical underpinnings of each model to other populations. The tenets of CBM—frequent, repeated measurement of key student outcome variables in an academic area for evaluating intervention effects—would be very useful for preschool populations.

For example, the Primary Prevention of Early Academic Problems (PPEAP) project currently is exploring downward extensions of CBM procedures to the kindergarten and first-grade levels (Good, Kaminski,

Schwarz, & Doyle, 1990). For preschool populations in particular, measures are needed that provide an estimate of the slope of pupil progress and a basis for ongoing, sequential decision making, with frequent opportunities to revise evaluations of risk (MacMann et al., 1989).

Use of the Measures in Decision Making

We propose that evaluation of the utility of CBA be conducted within the framework of the problem-solving decisions (e.g., Problem Identification, Evaluating Solutions) described earlier in this chapter. These decisions form one dimension of Figure 2. The second dimension is that of the specific school-aged population that is to be investigated, preschool, elementary, and secondary pupils. A third dimension is that of a particular CBA model.

Interpreting Figure 2 then, one can identify research questions in Problem Identification with elementary-aged students using Gickling's CBA-ID model or Evaluating Solutions with secondary-aged students using CBA-CR strategies.

Research on problem identification and certification. With elementary-aged pupils, we believe that research on the use of CBM strategies as a reliable method of problem identification and certification (Shinn, Tindal, & Stein, 1988) has been exhausted. No more studies are really needed to confirm that students placed in special education generally are the lowest performers in a curriculum compared to their local peers. Few, if any, problem-identification studies have been conducted at the secondary or preschool levels with CBM. No published studies have been conducted using other models of CBA for making these kinds of decisions. If problem identification continues to be seen as an area of priority (which, for the most part, we do not), then research using other models and populations other than elementary-aged students should be conducted.

Research on exploring solutions. The major use of nearly all CBA procedures has been on identifying the content of instructional interventions, the "what to teach" (Marston, 1989). The underlying premise is that better assessment data about what students can do and need to do will result in better learning. In a sense, then, CBA data are independent variables that should be demonstrated empirically to improve student outcomes. In many ways, the intervention-planning information provided by CBA is a treatment that can be tested by using a treatment-evaluation model (Deno, 1986). As just one example, Gickling and Thompson (1985) propose that if students are placed in

instructional-level material they will make progress. If students are not placed in instructional-level material (i.e., frustration- or independent-level material), they will not make as much progress. Although this conception has great intuitive appeal, we argue the need for data on the effects of Gickling's placement criteria and suggest that other criteria may work better. The contribution of instructional placement criteria using CBM strategies (e.g., Deno & Mirkin, 1977) also lends itself to empirical investigation. The types of interventions derived from CBA data are virtually limitless. Given the magnitude of instructional problems in schools, we believe great efforts are needed to determine how data can be used to increase the likelihood of implementing effective programs and decrease the likelihood of implementing ineffective programs.
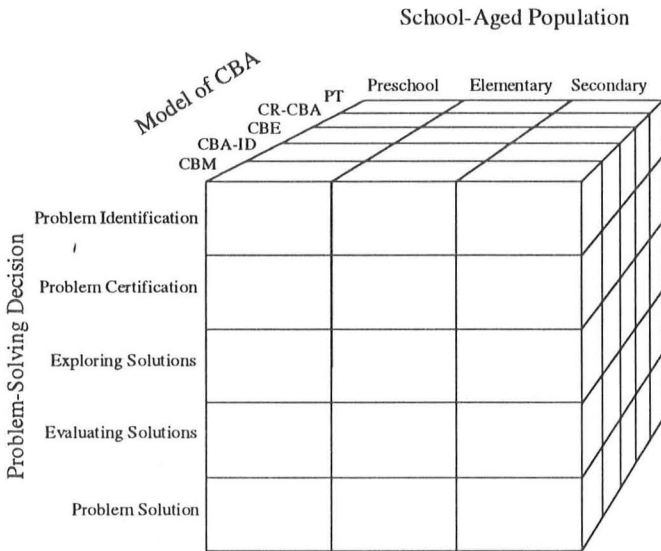
Figure 2. A matrix of research domains depicted by the type of problem-solving decision, type of CBA model, and target school-aged population.

A second key component of the Exploring Solutions decision is the specification of goals that are to be used to evaluate the effects of the intervention. Again, most of the research in using CBA to establish goals has been conducted within a CBM framework. The investigation of the effects of different goal structures and strategies on students' rates of progress and teaching (Fuchs & Fuchs, 1986; Fuchs, Fuchs, & Deno, 1985; Fuchs, Fuchs, & Hamlett, 1988; Fuchs, Hamlett, & Fuchs,

1988) has had fascinating outcomes. Some of the studies have been descriptive and need further experimental testing, however. For example, setting ambitious curricular goals has been associated with improved student outcomes (Fuchs et al., 1985). Other areas of research, such as the use of dynamic goals that change over time (Fuchs et al., 1988a), need replication.

Research on evaluating solutions. One of the most neglected decisions in schools is that of evaluating the effectiveness of interventions that are implemented. Far too often, *no* systematic data are collected to determine if what is implemented is working with individual students. When data are considered, they tend to be subjective opinions. Given the unique learning needs of individual students, as much or more assessment time and resources should be devoted to evaluating an intervention's effects as were used to identify the intervention's components. The evaluation of an intervention's effects using the curriculum in which students are instructed seems to be a logical process. Unfortunately, few *systematic* procedures for evaluating interventions using CBA have been specified in the professional literature. Even less research has been conducted in this area, with the exception of CBM. Within CBM, a host of research topics remain in making intervention effectiveness decisions. Among the important topics are further explication of the assets and liabilities of short-term versus long-term measurement with respect to estimating true progress, frequency of measurement, methods of summarizing student performance over time, and methods for increasing the frequency and effectiveness of changes in intervention strategies as a function of student performance data. Research on the use of computers in each of these areas (see Fuchs et al., 1988b, as well as this volume, for more details) also is increasing in prominence.

Efforts need to be increased on the use of other CBA strategies for evaluating interventions, in large part because CBM has been employed only to evaluate the effects of interventions in basic skill areas. Mastery monitoring approaches, where students' rates of progress through curricular objectives are examined (Jenkins, Deno, & Mirkin, 1979), remain potentially the most useful method in other curricular areas, especially for very young pupils and in secondary content areas. Unfortunately, mastery monitoring approaches have very few systematic procedures and virtually no research.

Problem solution. Problem solution decisions are made to determine if a problem is resolved and no longer requires additional resources. How do we know, for example, that an intervention has accomplished its purpose? In special education or Chapter I programs, this question

would be translated to mean, "What data do we have to suggest that special services are no longer required and a student may receive his/her instruction with other more typical students?" The use of data to make Problem Solution decisions is likely the least well-investigated area in education in general.

A problem-solving model would define a problem as resolved when the difference between what is expected and what occurs is no longer socially important. The use of student performance data in a curriculum again is logical for operationalizing what is expected and what is occurring and therefore may be useful in making this decision. No systematic procedures have been identified or developed, however. As a result, no empirical work has been accomplished, regardless of CBA model. Implementors of CBM (Allen, 1989; Shinn & Rodden-Nord, 1990) have begun a series of processes to assist educators in making Problem-Solution decisions.

Research on Implementation

Most research on CBA strategies has been microcosmic, how specific measurement techniques work, and with what effects or how teachers can use specific decision rules to determine when to change their instructional programs. Very little research has been undertaken at a more molar, systems level, investigating, for example, what factors expedite or impede implementation. To date, the research that has been conducted has been constrained to CBM and from a retrospective perspective (Deno & Marston, 1989). Efforts should be made to study systems' reactions to implementation during the process of changes in assessment practices.

School district leaders (e.g., Germann, 1987) have identified a series of steps that are purported to increase the ease of implementation of CBM. If CBA is seen as a potential technology that should be implemented, then it seems logical that research on implementation should be conducted to facilitate the technology transfer. Prevailing opinion is that widespread changes occur neither easily nor frequently in education (Baer & Bushell, 1981; Cuban, 1990). Resistance factors should be identified and addressed.

CBA approaches, independently or in combination, represent innovations that will require change(s) in how schools operate. The assessment practices of school psychologists and special educators can be expected to change, as will the way the various service consumers (e.g., parents, teachers) accept and use the information that is provided. With reduced time spent on problem-identification and certification

decisions, it will be important to examine whether there are shifts in time devoted to intervention planning and evaluation of outcomes, and whether intervention services and resources can be restructured to serve students more effectively.

CLOSING COMMENTS

CBA represents an important innovative assessment technology that has the potential to improve students' educational programs. We are pessimistic about whether the various CBA systems will be implemented with sufficient fidelity to improve outcomes, however. Although the appeal of using testing materials derived from students' curricula is obvious, we are of the opinion that the initial attraction may, in fact, be a distraction. That is, the use of content-valid tests is a necessary but not sufficient step for better educational assessment and decision-making practices. Just the use of content-valid tests stops at the superficial benefits of an alternative educational assessment approach. As we have illustrated, there is much more to improved educational assessment practices: A substantive shift in assessment paradigms is required. Through our examination of the literature written about CBA by its contributors and noncontributors, we believe that many knowledgeable persons are *not* seeing the required paradigmatic shift, and that what we will see is merely another test added to the repertoire of school psychologists and special educators. Better educational assessment practices cannot "combine state of the art regression discrepancy and curriculum-based models" (CASP, 1990, p. 12). Instructional plans derived from a profile analysis of WISC-R protocols are not well-wed to an analysis of CBA student error types.

Earlier, we pointed out Sarason's belief that school change comes when the system's values suggest that changes are necessary. We stated our own belief that leaders in school psychology have established a value system in which CBA may be integral. However, we are concerned that the "base of the triangle is not wide enough" to support the calls for changes in educational assessment practices espoused by CBA. That is, there may not be enough sufficiently trained personnel to implement quality educational assessment practices, including CBA, with sufficient integrity to change existing regularities. Training occurs at two levels, preservice and in-service. Bardon (1988) has pointed out the difficulties in training at both levels. The former requires training by institutions of higher education, which, as Bardon describes, are slow themselves to adopt new approaches. The difficulties of in-service training are compounded by the fact that many practitioners

consider themselves already trained and see little need for additional training, especially at the fundamental, conceptual level and to the degree that would be required by a major paradigm switch. For success, we will need to train well a *generation* of university trainers and school personnel. Changes in training programs may be occurring, but to date, changes in educational assessment training practices are not obvious (Reschly, Genshaft, & Binder, 1987).

Lest we close on a gloomy note, let us add that generally, schools that have implemented CBA-type procedures with integrity have reported positive outcomes (Germann & Tindal, 1985; Marston & Magnusson, 1985; 1988). Further, CBM is serving as an integral component of statewide adoption of a problem-solving assessment model and special education reform (Iowa State Department of Education, 1990).

In analyzing the characteristics of effectively implemented interventions described by Rogers (1983) (e.g., relative advantage, trialability, observability), we believe that each and all models of CBA possess many of these characteristics. The future of improved educational assessment using CBA strategies is filled with potential. We encourage a well-thought-out implementation process that exploits the limited technical assistance that is available.

REFERENCES

Alessi, G. (1988). Diagnosis diagnosed: A systemic reaction. *Professional School Psychology, 3*, 145-152.

Alessi, G. (1988). Direct observation methods for emotional/behavior problems. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Conceptual foundations and practical applications* (pp. 14-75). New York: Guilford.

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

Baer, D. M., & Bushell, D. (1981). The future of behavior analysis in the schools? Consider its recent past and then ask a different question. *School Psychology Review, 10*, 259-270.

Bagnato, S.J., Neisworth, J.T., & Munson, S.M. (1989). *Linking developmental assessment and early intervention: Curriculum-based presecriptions* (2nd ed.). Rockville, MD: Aspen.

Bailey, D. B. (1981). Investigation of learning measures as screening procedures with kindergartners. *Psychology in the Schools, 18*, 489-

495.

Bardon, J. I. (1988). Alternative educational delivery approaches: Implications for school psychology. In J. L. Graden, J. E. Zins, & M. C. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 563-571). Washington, DC: National Association of School Psychologists.

Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). *The scientist practitioner: Research and accountability in clinical and educational settings.* New York: Pergamon.

Barrios, B. A. (1988). On the changing nature of behavioral assessment. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook* (3rd ed.) (pp. 3-41). New York: Pergamon.

Blankenship, C. (1985). Using curriculum-based assessment data to make instructional decisions. *Exceptional Children, 52,* 233-238.

Bursick, W. D., & Lessen, E. (1987). A classroom-based model for assessing students with learning disabilities. *Learning Disabilities Focus, 3,* 17-29.

California Association of School Psychologists. (1990). *New directions: Registration packet for the California Association of School Psychologists 41st Annual Convention.* Millbrae, CA: Author.

Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher, 18* (1), 26-31.

Coates, R. (1989). Faint praise. *NASP Communique, 16,* 3. Washington, DC: National Association of School Psychologists.

Coulter, W. A. (1988). Curriculum-based assessment: What's in a name? *National Association of School Psychologists Communique, 16,* 13.

Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher, 19* (1), 3-13.

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

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. (1989). Curriculum-based measurement and alternative special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1-17). New York: Guilford.

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19,* 1-16.

Deno, S. L., & Marston, D. (1989). Evaluation of CBM in public schools: Grant continuation proposal, U.S. Department of Education.

Minneapolis, MN: University of Minnesota, Special Education.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual.* Reston, VA: Council for Exceptional Children.

Deno, S. L., Mirkin, P. K., & Shinn, M. R. (1979). *Behavioral perspectives on the assessment of learning disabled children* (Monograph No. 12). University of Minnesota, Institute for Research on Learning Disabilities.

Deno, S. L., Mirkin, P. K., & Wesson, C. L. (1984). How to write effective data-based IEPs. *Teaching Exceptional Children, 16,* 99-104.

Deno, S. L., Wesson, C., & King, R. P. (1984). Direct and frequent measurement of student performance: If it's so good for us, why don't we do it? *Learning Disability Quarterly, 7,* 45-48.

Fuchs, L. S. (1986). Monitoring progress among mildly handicapped pupils: Review of current practice and research. *Remedial and Special Education, 7,* 5-12.

Fuchs, L. S., Deno, S. L., & Marston, D. (1983). Improving the reliability of curriculum-based measures of academic skill for psychoeducational decisionmaking. *Diagnostique, 8,* 135-149.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1983). Data-based program modification: A continuous evaluation system with computer software to facilitate implementation. *Journal of Special Education Technology, 6,* 50-57.

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,* 449-460.

Fuchs, L. S., & Fuchs, D. (1984). Criterion-referenced assessment without measurement: How accurate for special education? *Remedial and Special Education, 5,* 29-32.

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

Fuchs, L. S., & Fuchs, D. (1986b). Curriculum-based assessment of progress towards long- and short-term goals. *The Journal of Special Education, 20,* 69-82.

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

Fuchs, L. S., Fuchs, D., & Deno, S. L. (1985). The importance of goal ambitiousness and goal mastery to student achievement. *Exceptional Children, 52,* 63-71.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1988). *Computer applications to*

curriculum-based measurements: Effects of teacher feedback systems. Nashville, TN: Peabody College, Vanderbilt University.

Fuchs, L. S., Fuchs, D., & Tindal, G. (1986). Effects of mastery learning procedures on student achievement. *Journal of Educational Research, 79,* 286-291.

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1988). *Improving data-based instruction through computer technology: Description of Year 3 software.* Nashville, TN: Vanderbilt University.

Galagan, J. E. (1985). Psychoeducational testing: Turn out the lights, the party's over. *Exceptional Children, 52,* 288-298.

Germann, G. (1987, November). *Factors in CBM administration.* Presentation at the Minneapolis Special Education Leadership Conference, Minneapolis, MN.

Germann, G., & Tindal, G. (1985). Applications of direct and repeated, curriculum-based assessment. *Exceptional Children, 51,* 110-121.

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

Gickling, E., & Thompson, V. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52,* 153-165.

Good, R., Kaminski, R., Schwarz, I., & Doyle, C. (1990, April). *Using pupil progress to identify at-risk kindergarten students: Pilot study.* Paper presented at the 22nd Annual Convention of the National Association of School Psychologists, San Francisco, CA.

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

Graden, J. L., Zins, J. E., & Curtis, M. J. (1988). *Alternative educational delivery systems: Enhancing instructional options for all students.* Washington, DC: National Association of School Psychologists.

Hall, G., & Hord, S. (1984). *Change in schools: Facilitating the process.* Albany, NY: State University of New York Press.

Hargis, C. H. (1987). *Curriculum-based assessment: A primer.* Springfield, IL: Charles C. Thomas.

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1986). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42,* 963-974.

Heller, K. A., Holtzman, W., & Messick, S. (1982). *Placing children in special education: A strategy for equity.* Washington, DC: National Academy Press.

Howell, K. (1986). Direct assessment of academic performance. *School Psychology Review, 15*, 324-335.

Howell, K. W., & Kaplan, J. S. (1980). *Diagnosing basic skills: A handbook for deciding what to teach*. Columbus, OH: Charles Merrill.

Howell, K.W., & Morehead, K. (1987). *Curriculum-based evaluation for special and remedial education*. Columbus, OH: Charles Merrill.

Idol, L., Nevin, A., & Paolucci-Whitcomb, P. (1986). *Models of curriculum-based assessment*. Rockville, MD: Aspen.

Iowa State Department of Education. (1990). *Problem-solving assessment series*: Des Moines, IA: Author.

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

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

Kratochwill, T. R., & Shapiro, E. S. (1988). Introduction: Conceptual foundations of behavioral assessment in schools. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Conceptual foundations and practical applications* (pp. 1-13). New York: Guilford.

Lentz, F. E. (1988). Direct observation and measurement of academic skills: A conceptual review. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Conceptual foundations and practical applications* (pp. 76-120). New York: Guilford.

Lentz, F. E., & Shapiro, E. S. (1986). Functional assessment of the academic environment. *School Psychology Review, 15*, 346-357.

Lombard, T. (1988a). Caution urged in embracing CBM assessment. *NASP Communique, 20*.

Lombard, T. (1988b). *Curriculum-based measurement: Megatesting or mctesting*. St. Paul, MN: Minnesota State Department of Education.

Macmann, G. M., Barnett, D. W., Lombard, T. J., Belton-Kocher, E., & Sharpe, M. N. (1989). On the actuarial classification of children: Fundamental studies of classification agreement. *The Journal of Special Education, 23*, 127-149.

Marston, D. (1988). The effectiveness of special education: A time-series analysis of reading performance in regular and special education classrooms. *The Journal of Special Education, 21*, 13-26.

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

Marston, D., Fuchs, L. S., & Deno, S. L. (1986). A comparison of standardized achievement tests and direct measurement techniques

in measuring student progress. *Diagnostique, 11,* 77-90.

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

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

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

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

National Association of School Psychologists/National Advocacy Coalition for Children. (1985). *Position statement: Advocacy for appropriate educational services for all children.* Washington, DC: Author.

Neisworth, J. T., & Bagnato, S. J. (1986). Curriculum-based developmental assessment: Congruence of testing and teaching. *School Psychology Review, 15,* 266-287.

Reschly, D. (1988). Special education reform: School psychology revolution. *School Psychology Review, 17,* 459-475.

Reschly, D. J., Genshaft, J., & Binder, M. S. (1987). *The 1986 NASP survey: Comparison of practitioners, NASP leadership, and university faculty on key issues.* Washington, DC: National Association of School Psychologists.

Reschly, D. J., Kicklighter, R., & McKee, P. (1988a). Recent placement litigation, Part I, Regular education grouping: Comparison of Marshall (1984, 1985) and Hobson (1967, 1969). *School Psychology Review, 17,* 9-21.

Reschly, D. J., Kicklighter, R., & McKee, P. (1988b). 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, 17,* 22-49.

Reschly, D. J., Kicklighter, R., & McKee, P. (1988c). Recent placement litigation, Part III, Analysis of differences in Larry P., Marshall, and S-1 and implications for future practices. *School Psychology Review, 17,* 39-50.

Rogers, E. (1983). *Diffusion of innovations* (3rd ed.). New York: Free Press.

Rosenfield, S., & Rubinson, F. (1985). Introducing curriculum-based

assessment through consultation. *Exceptional Children, 52,* 282-287.

Salvia, J., & Hughes, J. (1989). *Curriculum-based assessment.* New York: Prentice-Hall.

Salvia, J., & Ysseldyke, J. E. (1987). *Assessment in special and remedial education* (4th ed.). Boston: Houghton Mifflin.

Sarason, S. B. (1982). *The culture of the school and the problem of change* (2nd ed.). Boston: Allyn & Bacon.

Shapiro, E. S., & Lentz, F. E. (1985). Assessing academic behavior: A behavioral approach. *School Psychology Review, 14,* 325-338.

Shinn, M. R. (1986). Does anyone care what happens after the referral-test-placement process? The systematic evaluation of special education effectiveness. *School Psychology Review, 15,* 49-58.

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. (1989a). *Curriculum-based measurement: Assessing special children.* New York: Guilford.

Shinn, M. R. (1989b). Identifying and defining academic problems: CBM screening and eligibility procedures. In M. R. Shinn (Ed.), *Curriculum-based measurement: Asssessing special children* (pp. 90-129). New York: Guilford.

Shinn, M. R., Gleason, M. M., & Tindal, G. (1989). Varying the difficulty of testing materials: Implications for curriculum-based measurement. *The Journal of Special Education, 23,* 223-233.

Shinn, M. R., Good, R. H., & Stein, S. (1989). Summarizing trends in student achievement: A comparison of evaluative models. *School Psychology Review, 18*(3), 356-370.

Shinn, M. R., Goodwin, M., & Habedank, L. (1989). *Assessment of educational outcomes: Are we using the appropriate data to demonstrate effects?* Milwaukee, WI: Association of Applied Behavior Analysts.

Shinn, M. R., & Marston, D. (1985). Differentiating mildly handicapped, low-achieving and regular education students: A curriculum-based approach. *Remedial and Special Education, 6,* 31-45.

Shinn, M. R., Nolet, V., & Knutson, N. (1990). Best practices in curriculum-based measurement. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (2nd ed.. pp. 287-308). Washington, DC: National Association of School Psychologists.

Shinn, M. R., & Rodden-Nord, K. (1990). *A model of responsible reintegration of mildly handicapped students into general education classrooms.* Eugene, OR: University of Oregon.

Shinn, M. R., Rosenfield, S., & Knutson, N. (1989). Curriculum-based assessment: A comparison of models. *School Psychology Review, 18,*

299-315.

Shinn, M. R., Tindal, G., Spira, D., & Marston, D. (1987). Practice of learning disabilities as social policy. *Learning Disability Quarterly, 10*(1), 17-28.

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

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

Singer, J. D., Palfrey, J. S., Butler, J. A., & Walker, D. K. (1989). Variation in special education classification across school districts: How does where you live affect what you are labeled? *American Educational Research Journal, 26*, 261-281.

Taylor, R. L., Willits, P. P., & Richards, S. B. (1988). Curriculum-based assessment: Considerations and concerns. *Diagnostique, 14*, 14-21.

Tindal, G. A., Fuchs, L. S., Fuchs, D., Shinn, M. R., Deno, S. L., & Germann, G. (1985). Empirical validation of criterion referenced tests. *Journal of Educational Research, 78*, 203-209.

Tindal, G. A., & Parker, R. (1989). Development of written retell as a curriculum-based measurement in secondary programs. *School Psychology Review, 18*, 328-343.

Tindal, G. A., Shinn, M. R., & Germann, G. (1987). The effect of different metrics on interpretations of change in program evaluation. *Remedial and Special Education, 8*, 14-28.

Tucker, J. (1985). Curriculum-based assessment: An introduction. *Exceptional Children, 52*, 199-204.

Wesson, C. L. (1987). Increasing efficiency. *Teaching Exceptional Children, 20*, 46-47.

Wesson, C. L., King, R. P., & Deno, S. L. (1984). Direct and repeated measurement of student performance: If it's good for us, why don't we do it? *Learning Disability Quarterly, 7*, 45-48.

Will, M. (1986). *Educating students with learning problems: A shared responsibility.* Washington, DC: U.S. Department of Education, Office of Special Education and Rehabilitative Services.

Will, M. (1989). *The role of school psychology in providing services to all children.* Washington, DC: U.S. Department of Education, Office of Special Education and Rehabilitative Services.

Author Index

Subject Index