

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Curriculum-Based Measurement

Buros-Nebraska Series on Measurement and  
Testing

---

1993

## 2. A Review of Curriculum-Based Procedures on Nine Assessment Components

Gerald Tindal

University of Oregon, gerald@uoregon.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/buroscurriculum>



Part of the [Curriculum and Instruction Commons](#), and the [Educational Assessment, Evaluation, and Research Commons](#)

---

Tindal, Gerald, "2. A Review of Curriculum-Based Procedures on Nine Assessment Components" (1993). *Curriculum-Based Measurement*. 4.

<https://digitalcommons.unl.edu/buroscurriculum/4>

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Curriculum-Based Measurement by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# 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 comparison 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

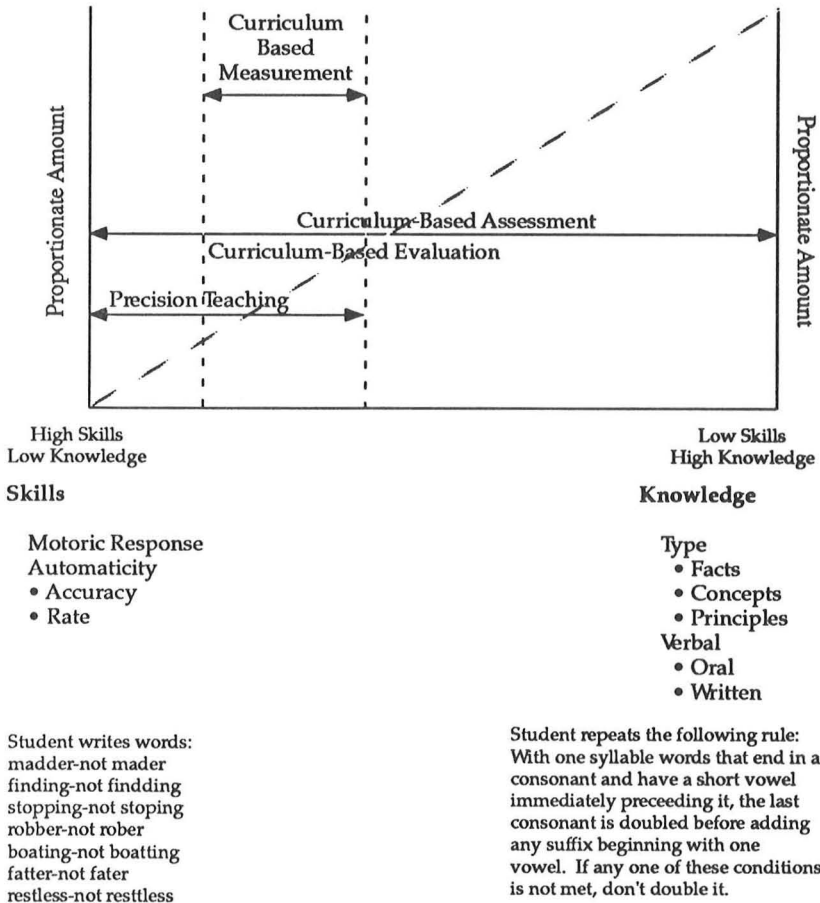


Figure 1. Skills versus knowledge focus of different models of curriculum-based procedures.

(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.

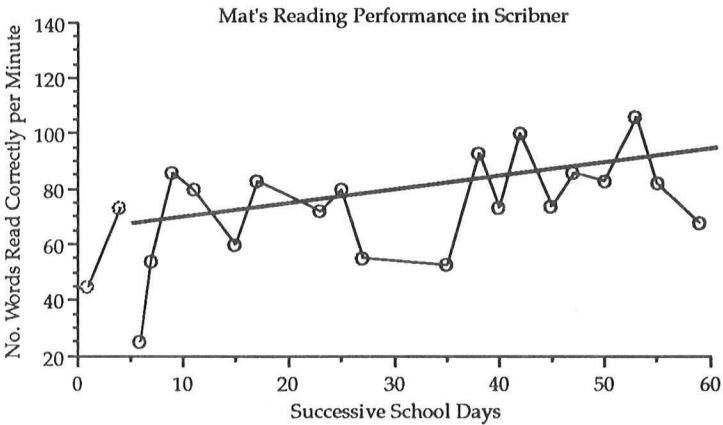
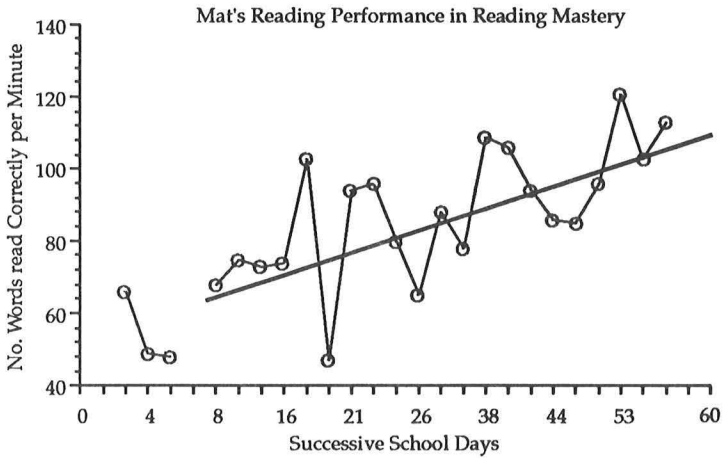


Figure 2. Student performance in the curriculum of instruction and the curriculum of the mainstream.

### 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 <del>lost</del>	Had had <del>(lost)</del>	He had <del>lost</del>	He had <del>lost</del>	He had <del>(lost)</del>	He had <del>lost</del>
Insertion	had I won	easily I won	big I won	should've I won	quickly I won	could've I won
Substitution	A big <del>party</del> pantry	The A big party	large A big party	busted A big party	The A big party	A big party <sup>1</sup> pantry
Mispronunciation	She'd <del>been</del> been out					She'd been out
Nonpronunciation	Rea <del>lly</del> bought <sup>1</sup>					
Aid/Assist		They <del>rested</del>	They (rested)	They <del>rested</del> <sup>2</sup>	They <del>rested</del> <sup>P</sup>	They <del>rested</del> <sup>P</sup>
Repetition	√ The latest	The latest <sup>1</sup>			The latest <sup>1</sup>	The latest <sup>R</sup>
Reversal		(Tired, old) feet			(Tired, old) feet	
Hesitation		Starved <del>for</del> food				Starved for food
<u>Non-Errors</u>						
Repetition			The latest <sup>R</sup>	The latest		
Self-Correct	He wasn't sure <sup>Sc</sup>	He wasn't sure <sup>sc</sup>	He wasn't sure <sup>Sc</sup>	He wasn't sure <sup>was</sup>		He wasn't sure <sup>S</sup>
Poor Phrasing	He ran home					He/ran/home
Hesitation			He ran home <sup>H</sup>	He ran home		He ran home <sup>✓</sup>
Partial Word					des- Destitute	
Non-words					He hadn't gone <sup>gone</sup>	
Punctuation		I won't. But she...				I won't. But she...
Dialect					He was talkin' loud	He was talkin' loud
Mispronunciation (Intonation)		Very plentiful <sup>M</sup>			Very plentiful <sup>✓</sup>	

<sup>1</sup> five seconds

<sup>1</sup> two or more

<sup>1</sup> Corrections vs. repetitions

<sup>1</sup> 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-efficient 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 comparison 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 research on graphic displays of student performance and data utilization, a number of issues have been addressed, including frequency of measurement (Mirkin, Deno, Tindal, & Kuehne, 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). ✓

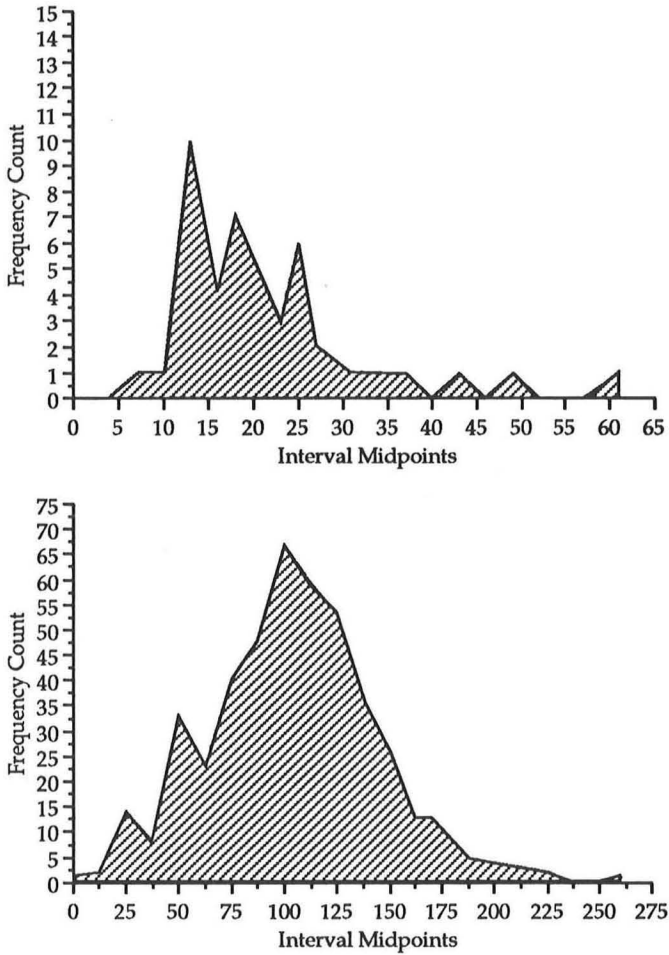


Figure 5. Graphic displays of data: Norm referenced distributions on oral reading fluency.

## 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-term goals to structure the outcomes (Fuchs & Fuchs, 1986; Tindal, Fuchs, Christenson, Mirkin, & Deno, 1981). These two procedures are illustrated in Figure 6 below:

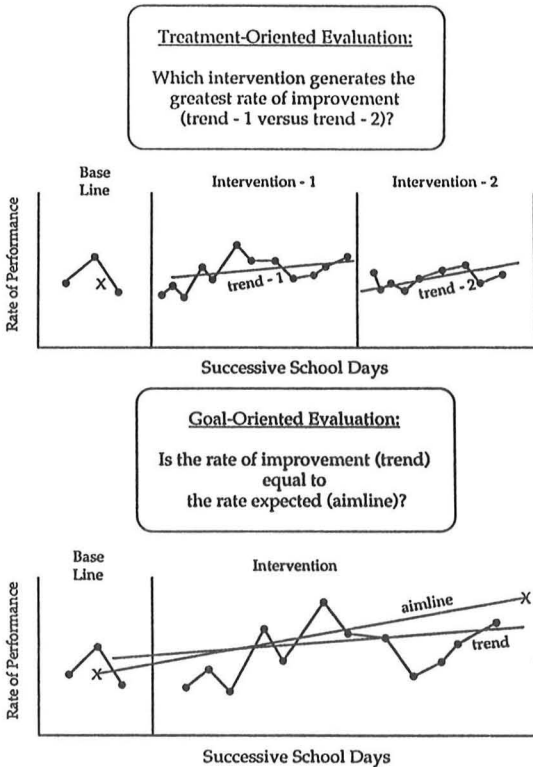


Figure 6. Two types of individual-referenced decisions: Treatment and goal oriented evaluation strategies.

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	
<b>Reference Type</b>	Norm-referenced		Criterion-referenced	Individual-referenced
<b>Curriculum Sampling</b>	Multi-curricula Multi-unit		Mono-curricula Mono-unit	Mono-curricula Multi-unit
<b>Data Scale</b>	Continuous		Discrete	Continuous
<b>Threats to Interpretation</b>	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.

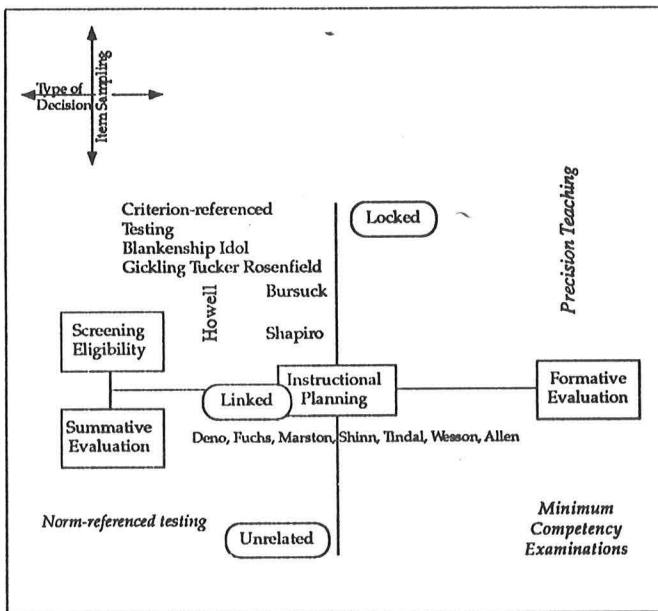


Figure 8. Comparison of different models of curriculum-based procedures on item sampling and type of educational decision.

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 comparison 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.