

Progress Monitoring for Specific Learning Disabilities (SLD) Eligibility Decisions

Introduction

Since the revised SLD rule went into effect in 2010, the Department has received a number of questions about the selection of progress monitoring tools used to collect data during intensive scientific, research-based or evidence-based interventions (SRBIs).

This document provides information and guidance in response to the specific question:

“How does a local educational agency determine whether to use a Curriculum Based Measure (CBM) or a Computer Adaptive Test (CAT) when collecting progress data during intensive scientific, research-based or evidence-based interventions for the purpose of making a special education eligibility decision in the area of Specific Learning Disabilities?”

Background

There are several types of progress monitoring tools that may or may not qualify as a “probe” as defined by the Wisconsin Specific Learning Disabilities (SLD) eligibility rule. Both Curriculum Based Measures (CBMs) and Computer Adaptive Tests (CATs) can be used in a school district’s [strategic assessment system](#) and can serve numerous purposes when making instructional decisions for all students. Both can be used for progress monitoring over relatively long periods of time (e.g., months, semesters, years). However, when considering an instrument to use to monitor progress over a short period of time — weeks or over the course of a special education evaluation timeline — local educational agencies (LEAs) will need to carefully consider the purpose for which the data are used; and the relative strengths and weaknesses of that measure to meet the specific purpose. This is particularly true when the data will be used to make a special education eligibility decision.

Local Decision

How a LEA conducts progress monitoring during intensive interventions is governed by administrative rule, the SLD rule (see [PI 11.02](#) and [PI 11.36\(6\)](#)). However, the choice of **which** progress monitoring probe to use when collecting progress data during a scientific, research- or evidence-based intensive intervention (SRBI) for the purpose of making a special education eligibility decision in the area of SLD is a local decision.

Decision Errors

Identifying a student as having a disability is a high stakes decision, both for the student and for the district. Decision errors are costly. Public schools spend an average of two to three times more money on each student eligible for special education as they do for students without disabilities (*Center for Special Education Finance*). Educational, employment and community-based outcomes for students with special educational needs are poor in comparison to students

educated entirely in the general educational system. Therefore, LEAs are keen to ensure they do not make the mistake of identifying a student as needing special educational services when s/he does not. More costly are the lifelong consequences that a student and her/his family bear when the student is misidentified as needing special education when s/he does not in fact have a disability.

There is a second type of decision error to consider: that of finding a student does not need special educational services when s/he does, in fact, need those services in order to progress in school. Such a student is at risk of not receiving the intensive interventions s/he needs to close the academic gap.

Threats to decision errors in SLD eligibility determinations can be reduced, among other means, by using valid and reliable data that are sensitive to growth over short periods of time.

Decision Guidelines Based on the Wisconsin Administrative Rule

Decisions are based on a variety of information including that from individually administered standardized tests of achievement, progress monitoring data collected during SRBIs, and observations during classroom instruction and intervention [PI 11.36(6)].

The Wisconsin Administrative Rule [Wis. Admin.Code § PI 11.02] offers the following definitions used during special education decisions about SLD eligibility:

- Progress monitoring is defined as “a scientifically based practice to assess pupil response to interventions” [Wis. Admin. Code § PI 11.02(10)]. Progress monitoring requires the use of scientifically based tools called probes to measure progress.
- Probes are
 - a. Brief,
 - b. Direct measures of specific academic skills,
 - c. With multiple equal or nearly equal forms,
 - d. That are sensitive to small changes in student performance and
 - e. That provide reliable and valid measures of pupil performance during intervention [Wis. Admin. Code § PI 11.02(9)].

The rule further requires Individualized Education Program (IEP) teams to analyze weekly progress monitoring data collected during two SRBIs implemented with fidelity in each area of concern. The IEP team analyzes the data to determine whether or not the student’s response to intensive intervention was sufficient or insufficient [Wis. Admin. Code § PI 11.36 (6)(c) 2.a.].¹

Progress Monitoring Probes

States and districts are in the early stages of knowing how to use CATs and CBMs for determining whether a student has or has not made sufficient academic progress. Each system has relative strengths and weaknesses; neither CBMs nor CATs function perfectly as probes. Those strengths and weaknesses may be considered tolerable when making relatively low stakes decisions. However, in a high stakes decision, such as when using a growth measure to

¹ Please note there are other provisions in the SLD rule that must be considered. This document focuses only on those provisions of the rule that relate to appropriate progress monitoring tools to use during intensive interventions that fulfill the requirements of the rule.

determine whether a student’s progress is sufficient or insufficient in the course of an evaluation used to determine eligibility for special education, the relative strengths and weaknesses are magnified. As such, LEAs have a duty to minimize the negative impact of high stakes decisions on students.

This table provides some explanation and detail on the relative strengths and weaknesses of these two methods of measurement, especially relative to purpose.

Comparison of CATs and CBMs for Use in SLD Identification²

Considerations	Computer Adaptive Tests (CATs)	Curriculum Based Measures (CBMs)
Theoretical Foundation	Item response theory (IRT): A mathematical model of the relationship between performance on a test item and the test taker’s level of performance on a scale of the proficiency being measured. In theory, the item’s difficulty level and discrimination (diagnostic accuracy) are controlled, thus accounting for error. With technological advances, there has been an acceleration of use of IRT in computer adaptive assessments.	Classical test theory: A psychometric theory based on the view that an individual’s observed score on a test is the sum of a true score component for the test taker plus an independent measurement error component. Reducing the size of the error is accomplished by quality test construction and implementation with fidelity. The Standard Error of Measurement for the score as well as for the slope (Standard Error of Estimate) may be calculated and analyzed by IEP teams.
Probe <i>a. Brief</i>	Each computerized assessment takes 20-30 minutes to administer; group administrations are possible; When administered individually, an adult is required to promote engagement and guard against technology failure. The amount of time needed for administration may be an important consideration since probes must be administered at least weekly when consistent with the SLD rule.	Assessments range from one minute to 20 minutes by type; Assessments are administered individually, with some exceptions (e.g., math problem-solving, correct written word sequences). Most probes require individual face-to-face administration by a trained adult.
Probe <i>b. Direct measure of specific academic skills</i>	CATs are not direct measures of specific academic skills and may not align with all targeted areas of concern listed in the SLD rule; items are not drawn from the same discrete skill area (e.g., items assessing basic reading skills are included along with items of reading comprehension). CATs tend to be considered measures of broad areas, such as reading comprehension or math problem solving despite the item content coming from other areas as well. CATs are available for two, or at most, three of the eight areas of concern. See [Wis. Admin. Code § PI 11.36 (6) 1].	CBMs are not direct curriculum measures; nor are they direct measures of specific academic skills; they are <i>indicators</i> of direct skills meaning that the CBM has high predictive validity of the direct skill. There are CBMs for some but not all targeted areas of concern listed in the SLD rule. CBMs are available for six of the eight areas of concern. See [Wis. Admin. Code § PI 11.36 (6) 1].
Probe <i>c. Has multiple equal, or nearly equal, forms</i>	Each item is equated using the Item Characteristic Curve from Item Response Theory after being field tested on students. Using item characteristic information, every	Probes are equated by exacting standards. Different publishers offer different numbers of probes in a probe set. Probes being equal or nearly equal is essential to reduce error. A

² This table is not an exhaustive list or comparison of measures, but rather, a guidepost meant to guide LEAs in planning the high stakes decisions surrounding SLD identification, a process governed by Wisconsin Administrative Rule.

	<p>item is placed on the test scale. This item by item equating process results in multiple equal or nearly equal forms. The pool of equated items is typically quite large. The item quality information is usually available in technical guidance from publishers.</p>	<p>larger number of unique, equated probes is useful when measuring weekly.</p>
<p>Probe <i>d. Sensitive to small changes in student performance (over weeks, for an SLD eligibility decision)</i></p>	<p>Currently available CATs have not achieved sensitivity to growth over weeks. Sensitivity to growth has been demonstrated over months, semesters or years. LEAs should review information for each test before determining whether it is appropriate for the particular use.</p>	<p>Each type of CBM has associated research indicating the sensitivity to demonstrate growth by weeks, and by age. Some CBMs have greater sensitivity than others (e.g., CBM-Reading vs. CBM-Writing). Sensitivity to growth is more robust at younger ages than older ages (e.g., late middle school or high school). LEAs should review information for each test before determining whether it is appropriate for the particular use.</p>
<p>Probe <i>e. Reliable and Valid</i></p>	<p>The Standard Error of Measurement (SEM), that is, the tendency of scores to vary because of random factors, is relatively high for CATs. This means that data sometimes may be difficult to interpret. For example, interpretation of slopes which are relatively flat is difficult, as one cannot distinguish whether the lack of growth is an artifact of the test or whether the student's rate of progress is poor.</p>	<p>Implementation integrity is required to achieve reliability of the probe score. Sufficient duration of weekly probe data is required to achieve reliability of the slope. No free, public domain CBM-Reading assessment is known to have achieved the level of reliability to meet this standard for reliability. Some educators perceive face validity to be low. For example, CBM-Reading Comprehension (e. g, Maze, CBM-Reading) probes are criticized for not truly measuring comprehension. Reliability and validity may vary by age and test type.</p>
<p>Scientifically-based practice</p>	<p>Peer reviewed research on item response theory and computer adaptive testing for screening has accumulated for decades. Research on CATs for progress monitoring is in early stages.</p>	<p>Peer reviewed research on CBMs for screening has accumulated for decades and for progress monitoring for about ten years.</p>
<p>Other factors</p>	<p>CATs require an adult to monitor computerized administration and to promote engagement.</p> <p>CATs can be used for instructional design and for information about the point of access in the sequence of the curriculum. They are strong screening tools. At this time, few commercial publishers market their CATs as appropriate for weekly progress monitoring. Their greatest vulnerability is sensitivity to growth over short periods of time.</p> <p>Another important vulnerability is mismatch between the tool and intervention target. Results about progress are meaningful for decision-making only if you actually intervened in area measured.</p> <p>Eligibility decisions must be made using a measurement tool which matches the target area of the intervention.</p>	<p>Requires adult one to one administration for most measures; group administration for some measures.</p> <p>CBMs can be used for instructional design, and for better understanding the nature of reading problems. They can be used for screening and frequent progress monitoring. Their greatest vulnerability is poor implementation integrity which results in poor reliability of the score and of the slope.</p> <p>Another important vulnerability is mismatch between the tool and intervention target. Results about progress are meaningful for decision-making only if you actually intervened in area measured.</p> <p>Eligibility decisions must be made using a measurement tool which matches the target area of the intervention.</p>

Summary

The Department promotes the use of a strategic assessment system, which encourages LEAs to use various assessment tools that are matched to a particular purpose. Few measures are totally successful for more than one purpose. When the purpose is to monitor student progress during intensive interventions over a relatively short period of time in a manner consistent with Wisconsin's SLD rule, LEAs can reduce the chances of decision errors. LEAs may do so by choosing a measure that is matched to the intended purpose; valid and reliable; sensitive to growth over the relatively short period of time inherent in a special education evaluation timeline; implemented with integrity, and which best meets the requirements of the Wisconsin Administrative Rule [Wis. Admin. Code § PI 11.36(6)].