aimsweb

Measuring General Outcomes: A Critical Component in Scientific and Practical Progress Monitoring Practices

Mark R. Shinn, PhD
Professor of School Psychology, National Louis University
aimsweb Consultant

JANUARY 2013

Since last August, I have been on a serious quest to lose 30 pounds. Because of my weight loss intervention, I've changed what I eat, how much I eat, when I eat, and how much I exercise. This intervention generates a lot of potential data to track if I want to know if this is working. I choose to record the number of calories in each and every food I eat against a daily target. I record the type of exercise I engage in and the number of minutes I spend as well as at least one measure of the intensity of the workout so I can gauge the number of calories I'm burning. I also occasionally record my heart rate and, when I bike, my cadence or the number of pedal rotations per minute. In other words, I can, and do, collect a lot of information to help me judge how I'm progressing. But the bottom line in all this intervention activity is the single most important thing I do every day; the same way; at the same time; in other words, under standard conditions. I step on the bathroom scale and weigh myself. The "bottom line" is this: if I'm not losing weight, all the other data I collect are not especially meaningful and, overall, my weight loss intervention needs to be changed.

In Figure 1, I show my "bottom line."

FIGURE 1: A weight chart showing Mark's progress toward his weight loss goal over time.

I've gotten close to my weight loss goal, but I made more progress early in the intervention program. Recently, my intervention has been less effective and I need to make a program modification if I am to lose the remaining 5 pounds. I can make this decision with confidence. I make my choices about what types of data to collect and use to evaluate my weight loss intervention based on my training and knowledge of scientifically based progress-monitoring practices. The "bottom line" for my decision making is tied to the type of progress monitoring known as general outcome measurement (GOM).

Teachers' Progress Monitoring Practices, and the Purpose of this Paper

Teachers, too, recognize the importance of monitoring progress so that they can evaluate student learning and the effects of their teaching. The range of their progress monitoring practices varies considerably. Teachers may ask individual students spontaneous questions orally during instruction such as "So, what happened next?" Teachers may ask standard questions of all students using written, short answer questions such as "Why was the Declaration of Independence essential to the colonies' efforts to separate from England?" Elementary teachers regularly ask students to write orally dictated spelling words or write and revise a short paper. Elementary or middle school teachers may grade mathematics homework for the percentage of problems answered correctly, or give a mid-quarter that is a cumulative evaluation of all the mathematics skills taught to date. Teachers may assess progress by using end-of-unit tests that accompany published series

or through computer-driven adaptive tests that gives students easier problems when they struggle or more difficult problems when they succeed. Most of these progress monitoring practices are closely tied to progress monitoring questions like "how am I doing today, or this week, or this term?" Few of these practices are "bottom line," that is, capable of measuring important progress relevant for judging intervention success.

Unfortunately, for too many teachers, their progress monitoring practices are not tied to extensive training about the science of progress monitoring. Their intentions are good and the progress monitoring data that are collected are useful for making some, but not all, progress decisions, especially those related to the "bottom line." My point is not to judge this lack of knowledge, because to me, this is a deficiency in their higher education training. Instead, my point is to set the stage to improve the knowledge about the science of progress monitoring practices in general, and contribute to a fuller understanding of GOM, an essential tool in the toolbox of effective instruction. I will do this by linking my ideas to common practices, using everyday language, and limiting academic references to the following list of key readings (for which I will present a brief synthesis at the end of this paper).

Key Readings on the Science of Progress Monitoring

- Espin, C. A., McMaster, K., Rose, S., & Wayman, M. (Eds.). (2012). *A measure of success: How Curriculum-Based Measurement has influenced education and learning.* Minneapolis, MN: University of Minnesota Press.
- Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*, 488–500.
- Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28*, 659–671.
- Shinn, M. R. (2010). Building a scientifically based data system for progress monitoring and universal screening across three tiers including RTI using Curriculum-Based Measurement. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model, including RTI* (pp. 259–293). Bethesda, MD: National Association of School Psychologists.

I also will answer a series of frequently asked questions about GOM at the end of the paper.

Two Families of Progress Monitoring Practices

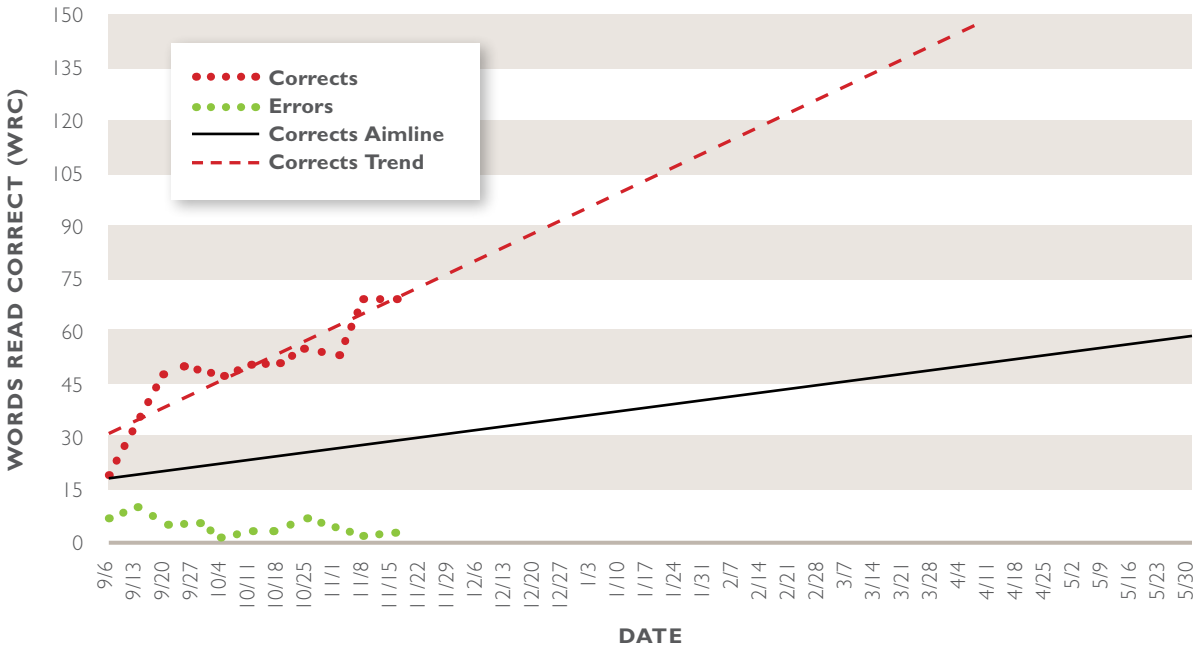
Within the science of progress monitoring, there are two major approaches or families: (a) general outcome measurement (GOM), sometimes known as long-term goal monitoring (LTG), and (b) mastery monitoring (MM), sometimes known as short-term goal monitoring (STM). Each approach is based on a fundamental set of assumptions with advantages and disadvantages that need to be understood to make valid decisions about student progress. The classic article comparing and contrasting these two families was published in 1991 by the noted progress monitoring scientists, Lynn S. Fuchs and Stanley L. Deno.

General Outcome Measurement. When I'm asked to describe the essential "big idea" of GOM, I like to say that it is a process of measuring one simple or "little" thing the same way over time to make a statement about something complex or "big." For example, for more than 30 years, educational research has demonstrated that teachers can measure something simple or "little" like oral reading for a short period of time (e.g., 1 minute) to make statements about something complex or "big" like general reading ability. Under standard conditions (e.g., "When I say begin…," timing for 1 minute, and counting the number of words read correctly (WRC)), educators can judge whether students are making reading progress or if a student is significantly discrepant

from other students in his or her reading skills. Much as I can answer the question(s) about my progress toward losing 30 pounds (i.e., am I losing the weight I want? Is my intervention working?), teachers can answer the important question, "Is this student becoming a better reader (or better in mathematics computation, or other basic skill area) or do I need to change my intervention?"

An example of an **aims**web frequent progress monitoring graph is shown in Figure 2.

FIGURE 2: A Grade 2 student's actual rate of progress compared to the expected rate of progress that reduces the reading achievement gap.



The black line is the rate of improvement (ROI) that would be required for this student to reduce the reading gap from peers. This expected rate of progress is called the *aimline*. Each red dot is a single data point obtained by having the student read a randomly selected Grade 2 reading passage from a pool of **aims**web Grade 2 reading passages. Each passage is of approximately equal difficulty. Each data point represents no more than 2 minutes of testing time under typical conditions. WRC is counted. The green dots are student errors. The red line is the *trendline*, or the actual rate of progress. The "bottom line" for this student is that the reading intervention is having a powerful effect. The student is improving much more rapidly than expected and is significantly reducing the achievement gap.

In other professions and disciplines, GOM is standard practice. For example, major decisions about the state of the national economy (big things) are made based on little things like the Dow Jones Industrial Average (DJIA), even though what is "measured" is based only on the average stock price of 100 of this country's thousands of businesses. Stocks are evaluated on their earnings per share. Major health decisions are made by measuring blood pressure or, in the case of a disease like diabetes, blood glucose levels. Cars generate different miles per gallon and housing-price changes are evaluated using the Shiller-Case Home Price Index.

If one feature stands out about GOM, it is its simplicity. Usually the measures are not time-consuming to collect, and they are collected the same way (i.e., the testing content and materials don't change each time). The data, once collected, are not difficult to organize, report, or understand. This feature leads to one of the major advantages of GOM, its feasibility. In other words, progress decisions can be made economically in terms of

time, cost, and complexity. I stand on my bathroom scale at the same time and under the same conditions and look at the number of pounds shown on the scale. A student whose reading progress is monitored weekly reads a randomly sampled reading passage of approximately the same difficulty (e.g., Grade 5) for 1 minute and the number of WRC is counted. Although the student may be taught using different instructional materials, of varying difficulty, over time, the progress monitoring tasks are the same in terms of difficulty and do not change. Therefore, change in performance is easy to see. When I stand on the weight scale, it doesn't change even though my weight-loss intervention may have changed. The weight scale doesn't become more, or less, sensitive to the bottom line.

The second major feature of GOM is that the test materials do not come from specific intervention materials or curriculum. Although the original research from which **aims**web testing practices were derived was built on using assessment materials from specific curricula, more than 20 years of accumulated evidence has demonstrated the improvement in progress monitoring using high quality test materials representing general curriculum. The **aims**web test materials across achievement domains (e.g., reading, mathematics) correspond to a general reading curriculum or a general mathematics curriculum. This feature leads directly to another advantage of GOM: progress monitoring is instructionally eclectic. Much as my weight scale works with any weight-loss program, educators can monitor the progress of their intervention the same way, whether the intervention is based on Program A for 90 minutes a day, Program B for 45 minutes a day, or Program A for 90 minutes followed by Program B for an additional 45 minutes.

The third and final major feature of GOM is that by using test materials of equal difficulty from general curriculum administered over time, students will be tested on content they have learned and content they will learn in the future. This features leads to an advantage of continuously testing for retention (i.e., maintaining what has been learned) and generalization (i.e., applying skills and strategies to untaught content). Of course, these two variables are common areas of concern; students often are reported to fail to retain, or fail to generalize, their learning.

But the real bottom-line advantage of GOM is that it is highly related to gains in student achievement. Frequent formative evaluation, consisting of progress monitoring during instruction on standard tasks of equal difficulty, results in some of the most powerful intervention effects in educational science. The reasons are clear: it makes little sense to keep delivering the same intervention when there are data showing no progress. When teachers have accurate information that validates student progress, an intervention is continued. When teachers have accurate information about lack of progress, the intervention can be changed.

Of course, GOM is not a perfect solution, and it requires test materials that meet certain criteria. The most challenging is that a general outcome measure be empirically validated as an indicator, or predictor, of proficiency in a broader domain. Something "little" or simple must be related to, or predictive of, something "big" or more complex. Fortunately, in the basic skill areas, a sizable amount of research on quality indicators has been accomplished and incorporated in **aims**web. For example:

In reading, an oral reading test of 1 minute where students read passages aloud yields a score (i.e., number of words read correctly) that is a valid indicator of general reading achievement. This indicator is called Reading-Curriculum Based Measurement (R-CBM).

In reading, a silent reading test of 3 minutes where students read a passage from which every seventh word is deleted, and fill each gap by selecting the one word (out of three choices) that preserves the meaning, is a valid indicator of general reading achievement for older students. This indicator is called Maze.

In mathematics, requiring students to write answers to a range of number-sense and computational problems for 8 minutes and counting the number of items answered correctly (with extra points for more difficult items) produces a score that is a valid indicator of general mathematics computation proficiency. This indicator is called Mathematics Computation (M-COMP).

In mathematics, requiring students to write answers to a range of word, concepts, and application problems for 8–10 minutes and counting the number of items answered correctly (with extra points for more difficult items) yields a score that is a valid indicator of general mathematics application and problem solving proficiency. This indicator is called Mathematics Concepts and Applications (M-CAP).

In basic skills written expression, requiring students to write a short essay for 3 minutes about a given topic and counting the number of words or correct writing sequences is a valid indicator of general writing skill. This indicator is called Written Expression CBM (WE-CBM).

In spelling, requiring students to write orally dictated words for 2 minutes and counting the number of correct letter sequences is a valid indicator of general spelling skill. This indicator is called Spelling CBM (S-CBM).

To date, using GOM progress monitoring with confidence is an emerging science in academic areas outside of these basic skills, and within the basic skills for typically developing students beyond Grade 8. Research programs in content areas such as history, science, and middle school and high school language arts, and with very young children, have shown potential in identifying valid indicators, but few approach the level where they can be said to be comparable with existing basic skill (e.g., CBM) measures.

Mastery Measurement. When I'm asked to describe the essential "big idea" of MM, I like to say that it is a process of measuring different things in different ways at different times to make statements about simple or "little" things. When teachers use MM while teaching students addition facts through 10 or -/oat/ combination words like /boat/, /coat/, /float/, they test students on exactly those skills and nothing else. Conclusions can be drawn only about whether a student has learned addition facts through 10 or -/oat/ combination words. Concluding that because students have learned addition facts they have become "better at mathematics," or that learning -/oat/ combination words makes them "better readers," would require making a higher-level inference that seems plausible but is often not correct.

How so? Let's go back to my weight-loss program. I am also counting calories daily relative to a daily calorie target. It is reasonable to assume that if I make my daily calorie target, I will be losing weight. But, for a variety of reasons, I've made my daily calorie target for a number of days and not lost weight.

Now, the primary advantage of MM is that it conveys important information to teachers about the immediate impact of an intervention, answering the question "did the student learn what I'm teaching" on a specific time frame (e.g., today, this week, this quarter). This information has high instructional validity. But instructional validity, as I suggested, is not often the best way to judge progress over the long haul. For example, a student may pass an end-of-unit addition facts test and move to the next unit on subtraction facts that is also then tested to evaluate learning. However, once a student passes the addition facts test, they may not systematically be evaluated on that content again; a student's inability to retain their addition facts might be missed.

Again, strong instructional validity is not irrelevant, and the feedback from MM is important for teachers in the near term. But we want to gauge progress toward larger domains or constructs rather than just specific skills.

The advantages of strong instructional validity, and the corresponding teacher comfort level, are offset by a number of disadvantages. First and foremost, when MM is done well, it can be logistically challenging and

therefore less feasible than GOM. In theory, a skill within a validated (i.e., proven to be effective) instructional sequence is measured repeatedly and frequently until the student demonstrates mastery, also using a criterion that is validated. Then, the student moves to the next skill(s) taught in the instructional sequence and the process is repeated. An illustration of a prototypical MM graph is shown in Figure 3.

FIGURE 3: A prototypical Mastery Monitoring progress monitoring graph showing repeated assessments within a skill area until mastery, then moving to the next skill taught.

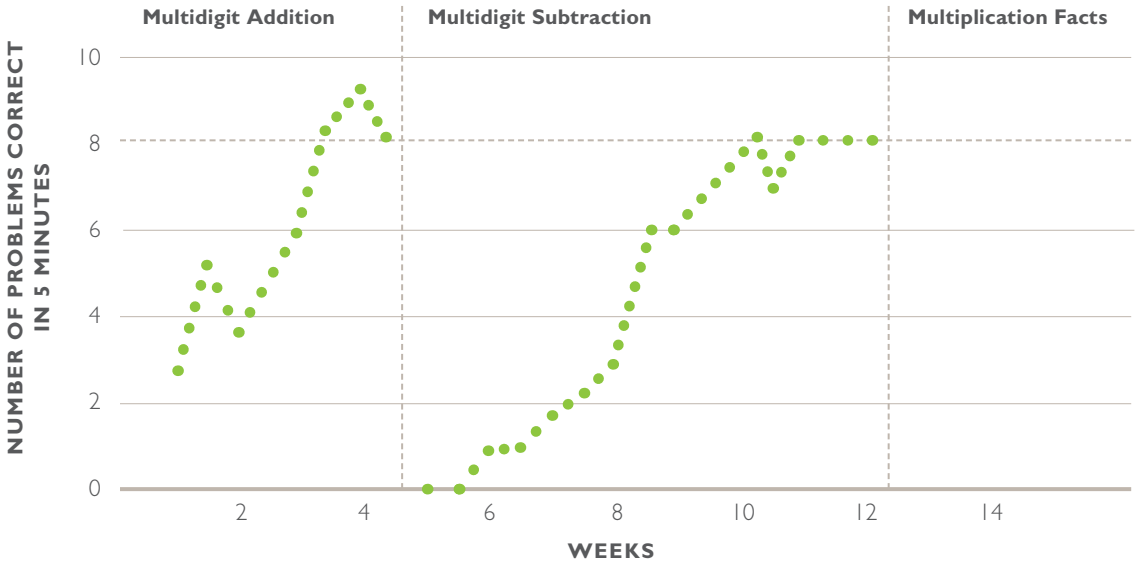


Figure adapted from Whitney Donaldson http://www.studentprogress.org/library/presentations.asp

Quality MM requires that for each skill or sets of skills, (a) reliable and valid tests are developed, and (b) there are multiple forms. This requirement, in and of itself, can be daunting, unless the tests are provided by the intervention program. Potentially compounding complexity is the ramifications of having each student continue to receive instruction in a skill until achieving mastery of that skill. The consequence is that unless all students achieve mastery at the same rate, instruction must be individualized so that Student 1 is in Skill A, while Student 2 is in Skill C, Student 3 in Skill D, etc. For teachers, this means trying to keep track of which students are being taught which skill(s), and which students need to be monitored on which corresponding tests.

One of the unintended consequences of MM is that because it can be logistically complicated, teachers often don't use the progress monitoring data they collect. That is, when student progress is monitored at the end of the instructional timeframe (e.g., a week), students move to the next skill(s) regardless of performance.

Perhaps not surprisingly, there is little compelling evidence that MM is related to meaningful improvement in student learning. I would argue that this lack of empirical support is not because MM cannot contribute to teachers' decisions about progress and, thus, the quality of their decisions about intervention effectiveness, but that because of its logistical complexity, MM is implemented very poorly.

Complementary, Not Competitive

In the important question of how to implement progress monitoring, it is important to recognize that it is not an either/or question with respect to GOM or MM. In fact, they are complementary, but not equal. The information they provide serves different purposes. GOM contributes to higher-stakes decisions such as "Is this student becoming a much better reader?" or "Is my Tier 3 intervention producing the results I need in math?" Additionally, GOM provides information about whether students are retaining and generalizing critical skills in basic skills areas. MM contributes to my understanding of the more immediate, but smaller stakes, instructional decisions regarding specific skills, and importantly can help me when there is less than satisfactory progress. If I use my weight loss intervention as an example, now that my progress has slowed, I need to look at my calorie information and/or exercise data to decide how to make intervention changes. If a choice must be made between one approach and the other, my scientific and practical recommendation would be to build a high quality GOM progress monitoring system in the area where we know how to do that, basic skills, and then use MM to support instructional decision making.

Summary

This paper was written to communicate some big ideas about General Outcome Measurement (GOM) and how it compares to another approach for progress monitoring, Mastery Measurement (MM). Certainly, the kinds of progress monitoring practices that most teachers use look more like MM, although I would argue that educators' implementations of MM often do not meet many of the requirements for good-quality mastery measurement. Progress monitoring practices instead tend to be more summative, occurring after instruction for the purpose of drawing conclusions, rather than formative, occurring during instruction for the purpose of judging intervention effectiveness and the need to modify instruction. Teachers value, as I do, instructional validity, and in some content areas (e.g., science, social studies, and advanced language arts or mathematics) or early childhood, MM may be our only choice due to the lack of validated GOM indicators. But in the basic skills areas, GOM is not only feasible, but critical in improving decision making regarding intervention planning and delivery. It addresses the "bottom line."

Frequently Asked Questions (FAQs)

Won't students get frustrated or discouraged if you are not testing them exactly on what they are learning now?

Perhaps the single biggest barrier to teachers' use of GOM is that students are not tested directly on what they are being taught. Relatedly, I hear teachers report that students will feel bad when they aren't making progress. I see these as two separate questions. Let me address the last first. We as educators should feel bad when students aren't making progress, but more importantly, we should change our intervention when we see a lack of meaningful progress. Students need to know that "we know" and are doing something about that lack of progress. Regarding the first question, I would argue that this has to do with teachers' confidence in a new progress monitoring tool. Will students improve when I test them this way? It is important to know that when GOM is used as the progress monitoring approach, students are expected to make slower progress than when MM is used, but there will be steady progress toward a long-term goal. If I return to my weight loss program metaphor, I couldn't expect to lose 30 pounds immediately, and the first few weeks were both daunting and moderately discouraging. But the scale did move in the desired direction! Additionally, many teachers do have more immediate information on the instructional validity of their outcomes. Although it might not be at the level of quality that good MM requires, teachers can and should use their existing data to augment decision making.

Shouldn't progress monitoring test students on exactly what they are learning?

Of course. Again, this is a question of instructional validity, but it is not the same as the bottom line decision about meaningful student progress. If I want to answer the question "Can the student read multi-syllabic words ending in –ingly?" I need to understand that that is a different question from "is this student becoming a better reader?" Most teachers have information, albeit informal, about student performance on what teachers are teaching. What we lack too often is information on the "bottom line" of meaningful progress.

Shouldn't achievement tests be related to state standards or the Common Core State Standards?

Different tests have different purposes. To me, all tests should measure important things, and if they are doing so, the tests should be related to one another. A test that measures a state's reading standards (or Common Core State Standards) should be highly related to another reading test if both tests are valid. But the question is, should they be related or should they be exactly the same? Any standards are a set of beliefs about what is important, presumably tied to science. However, currently, different states have different standards and the current Common Core State Standards are different from previous national standards. **aims**web is designed to measure general, broad, and important language arts (reading, spelling, written expression) and mathematics (number sense, computation, application and problem solving) constructs. **aims**web measures have been validated against a number of criterion measures of these constructs, including tests built on state standards. In other words, the measures have criterion-related (concurrent and predictive) validity. Other tests measure those standards directly. Their use leads directly to conclusions about attainment of those standards. **aims**web measures assess those standards indirectly. They can be used to predict proficiency on standards, but their more important use is for screening and especially progress monitoring than for program evaluation and accountability. The types of tests that directly measure standards are not designed for progress monitoring and especially frequent progress monitoring, and that is **aims**web's strength.

Foundational Articles for Understanding General Outcome Measurement

- Espin, C. A., McMaster, K., Rose, S., & Wayman, M. (Eds.). (2012). *A measure of success: How Curriculum-Based Measurement has influenced education and learning.* Minneapolis, MN: University of Minnesota Press.

  This edited book details the history and scope of Curriculum-Based Measurement (CBM), the primary tools for progress monitoring and screening in **aims**web. Noted international authorities contribute chapters detailing advances in knowledge and practice across a variety of ages, academic areas, and countries.

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

  This scholarly article by noted progress monitoring authorities describes the critical assumptions and features of General Outcome Measurement and Mastery Monitoring. Advantages and disadvantages of each approach are presented.

- Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28*, 659–671.

  This scholarly article updates the Fuchs and Deno critical assumptions and presents the features of high quality progress monitoring practices. These features then are illustrated by comparing three different ways of progress monitoring, including General Outcome Measurement and Mastery Monitoring using common practice tools.

- Shinn, M. R. (2010). Building a scientifically based data system for progress monitoring and universal screening across three tiers including RTI using Curriculum-Based Measurement. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model, including RTI* (pp. 259–293). Bethesda, MD: National Association of School Psychologists.

  This book describes an approach for schools to use in building their basic skills data system for purposes of progress monitoring across multi-tiered services and screening.

ABOUT THE AUTHOR

Mark R. Shinn, PhD, **aims**web Consultant, is currently a Professor of School Psychology at National Louis University in Skokie, IL. He spent 19 years as Professor of School Psychology and Professor of Special Education at the University of Oregon. Dr. Shinn has edited two books on Curriculum-Based Measurement (CBM) and co-edited three editions of *Interventions for Achievement and Behavior Problems*, including the recently released third edition entitled *Interventions for Achievement and Behavior Problems in a 3-Tier Model, including RTI* published in 2010 by the National Association of School Psychologists (NASP). He has published more than 90 refereed journal articles and book chapters on progress monitoring, screening, and identifying academic problems. From 2003–2007, Dr. Shinn was one of six members of the U.S. Department of Education, Office of Special Education Programs (OSEP) Technical Review Panel members of the National Center for Student Progress Monitoring (NCSPM).