

**Answers to Questions Posed During Webinar 2:
General Outcome Measurement and Mastery Monitoring
Mark R. Shinn, Ph.D.**

QUESTIONS ABOUT GENERAL OUTCOME MEASUREMENT AND MASTERY MONITORING

- **We are using AIMSweb for math and R-CBM. The math teachers want to drop it for next year because the students are assessed at the beginning of the year on skills that have not yet been taught. They think this is unfair to the children. In addition, they say the assessment does not match the curriculum. How can we answer the teachers concerns?**
- **What are some recommendations for when teachers report that progress on GOM progress monitoring improvement is not reflected in the classroom?**

This is a tough situation(s) and one that I tried to explain in the webinar and the white paper. Without trying to be critical, teachers are making judgments about something important to them: *content validity* and *instructional utility*. Every day, they wonder “are the students learning what I taught today (or this week)?” This is what I would describe as a “performance issue.” Are student doing what I taught them? Important for immediate instruction, but it is not the same question(s) asked when we are doing general outcome measurement, which is more about *progress* than performance. This is evaluated using *construct validity* and *consequential validity*. These two judgments are technical and are not part of teachers’ assessment training. Imagine someone judging how a blood pressure cuff could be used to make judgments about our overall health. If some started doing this to you or expected you to do it and make decisions without training and support, you might react with skepticism. Again, unfortunately, intensive assessment training is not part of our pre-service training and it is a scientific issue that we need to be relentless in addressing. In sum, teachers *are* and need to be, invested in assessing *performance*. But other tests and methods, standardized and non-standard, can do that. We *all* need to be invested in *progress*. And that is more than just about content validity.

- **I have teachers that do a SLA and find that the student is many years below in math using MCAP (7th grade with instructional level at 5th grade). Their intervention in a pre-algebra intervention skills in middle school. 1) They want to only use their MM/curriculum embedded PM tools as they feel it is more reflective of instruction and sensitive to progress. 2) They diminish role that these skills skill need attention, 3) they say there are not enough items to be sensitive to their intervention. How would you respond? Be glad to get my question answered later.**

Seems like there are two questions here, one implicitly about instruction, the other about GOM and MM. Let me address the progress monitoring one first. Of course teachers are correct! They should *also* use the curriculum-embedded PM tools. Again, MM and GOM is not “either-or.” Doesn’t it make sense to judge student performance (MM) and progress (GOM)? I would expect *rapid growth* in performance, given effective instruction, and *gradual growth* in progress as measured by GOM.

With respect to instruction, I’m not clear that the intervention program is targeted at their current instructional skill level. If students are significantly discrepant, if grade-level instructional materials are used for intervention, I would be concerned. Too often, I’ve seen mathematics interventions turn out to be a variation of “louder and longer.” That is, no specialized materials nor instructional strategies, often with very difficult and poorly designed general education mathematics curriculum.

QUESTIONS ABOUT TECHNICAL FEATURES OF PROGRESS MONITORING

- **As it appears that AIMSweb utilizes the methodology of single-subject design during the intervention phase, would you explain, either here or via a reference that I may read, as to the theory behind why AIMSweb only requires the median data point prior to intervention? My experience has been that the same rules apply in baseline as in intervention (i.e., level, trend, multiple measures over time - at least 3 data points, etc.).**

Conventionally, the median is the “least biased estimator of central tendency” when there is a small sample size and/or non-normal distribution. It is also very easy to compute.

- **How often should we be doing our progress monitoring? Is it more effective to PM weekly, bi-weekly, or more often than that? Should we PM more frequently for some subjects like reading more than math?**
- **What is your definition of frequent progress monitoring in tier 1 and 2?**

Let’s start “backwards.” For IEP goals, I would suggest 1-2 independent samples per week. For Tier 3, I would recommend 1 sample per week. For Tier 2, it is desirable to have 1 sample per week, but with students with less severe achievement needs, one might instead repeat the Benchmark Assessment (3 samples) once per month. This generates fewer data points, but also produces a reliable (high quality) data point. Tier 2 frequency is largely about the amount of time and resources available.

My bias at Tier 1 is 3-4 times per year, although I would advocate for monthly Benchmark Assessment at K and 1 if I had the commitment from teachers and staff. To me, this really communicates a strong commitment to getting students off to a healthy start.

- **What is the suggested number of data points for Tier II and III?**

This is a complicated question and it has to do with the amount of variability in the data as well as the slope. If all the data points are *below* the Aim Line and you have at least 5-7 of them, I would argue you know the intervention is not working, even if the data are highly variable. If all of your data points are above your Aim Line, I would maintain my intervention and question the goal. If it is somewhere in between, with reasonable variability—and here is where a confidence interval is helpful—the rule is that “more is better” but we can’t wait forever. I recommend a minimum of 10 data points, but that’s to *guide* thinking, not replace it.

QUESTIONS ABOUT ASSESSMENT AND DECISION MAKING

- **Can we get a bit more clarity regarding the difference between, say, R-CBM as a screening assessment (that would lead to diagnosis and intervention) as opposed to a formative assessment (that would give information on the kind of instruction needed)?**

The big difference is the number of samples collected in R-CBM and the frequency of data collection. For screening, because of potential passage differences, including student interest, for a relatively high stakes decision, we take the middle score (the median) of 3 *at a single point in time*. For frequent progress monitoring, we typically take *1 single sample, but repeat the process (using a different passage) over time*.

- **So as a 2nd grade teacher? What should my day look like? Should I be using Aimsweb progress monitoring with all my students for reading and math? If yes, where do I find those resources? And, do I need to be trained to do that type of testing?**

As a second-grade teacher, I would expect that you participate in Benchmark Assessment of your students, to me, a *shared* responsibility among all staff members. I would expect that you participate in monthly grade-level team meetings and contribute to *using* the Benchmark information to universally screen for basic skill discrepancies and triage (i.e., align students to intensity of treatment by severity of their discrepancy) to Tiers 2 or 3. I would expect that you understand the Benchmark Assessment results and are taught and supported to communicate the results to parents. I would encourage you to contribute to developing Literacy Portfolios for your students with video “snapshots” by video recording of their Benchmark R-CBM oral reading samples and pairing it with an AIMSweb graph of performance to share. I would expect that at your monthly grade-level meetings, you would *review* the progress of your students who are receiving Tier 2 intervention in particular. In sum, yes, you should participate in progress monitoring of all your students, but as part of a team. For reading, this is an estimated 15 minutes per student, per *year*. For mathematics at Grade 2, the tests are group administered and are 8 minutes, which, conducted three times per year is less than 30 minutes, not counting scoring time.

QUESTIONS ABOUT CBM, AIMSWEB, AND OTHER TESTS

- **Can you give some examples of MM and GOM? Such as NWEA or AimsWeb- what category do they fall in?**

These questions all have similar themes and they go back to the 4 purposes of assessment. Within our field, there are generally four held purposes of assessment,

1. Screening/identification,
2. Diagnosis/instructional planning,
3. Progress monitoring, and
4. Program evaluation/accountability.

NWEP MAP, to me, is best used as an accountability/program evaluation tool. Its adaptive testing model allows a reliable and valid mechanism of scaling achievement. For progress monitoring purposes, I believe it is more closely aligned with summative assessment rather than formative assessment. It is not designed to be administered frequently, which is one point of similarity between GOM and MM. At best NWEA MAP is 3-4 times per year. AIMSweb, on the other hand, organizes and reports CBM data and is particularly well suited for *frequent* progress monitoring that is used formatively. This frequency of progress monitoring is crucial for high quality Tier 2 and 3 progress monitoring, for monitoring progress toward IEPs, and if progress monitoring is to be part of specific learning disabilities

(SLD) eligibility in Response to Intervention (RtI). NWEA is both too infrequent for these purposes and also has not been validated to do so.

It is plausible that NWEA/MAP could be used as a Tier 1 progress-monitoring tool as well as for accountability/ program evaluation. It also could be used a “first pass” screener. I like, but have concerns about computer-based testing and accurate decisions for some students. Computerized tests assume students are appropriately motivated and focused and for lack of a better term, I find that sometimes kids are “kids” and not (highly) motivated adults. That is why my bias is that we “see” students perform; I like as authentic tests as possible as long as they are time efficient and why I especially prefer R-CBM and with young children in particular.

As per using NWEA/MAP for screening, with older students, I may use the data as the first step in a multiple-gating screening. That is, I might start with using my MAP data to identify those students I would screen individually using R-CBM.

Most, but not all of the AIMSweb measures would fit the definition of GOM. In the field of education, there are *few validated general outcome measures* and those that are, are in the basic skills of reading, mathematics, and written language, including spelling. For young students, I believe there is sufficient information to suggest that Letter Sounds is a good GOM. The lack of validated GOM tests is one feature that makes CBM unique. Examples of MM tests include end-of-unit tests, chapter tests, quizzes, specific skill tests, etc. I would also add that I consider LN to not be a PM tool, and PSF and NSF MM instruments.

- **I talk to a lot of schools who are trying to choose between AIMSweb GOMs and NWEA MAP. Any comments for them to help them choose?**
- **My district is proposing doing away with AIMSweb and replacing it with Measures of Academic Progress (MAP). I don't have a strong understanding of MAP yet but I don't believe it lends itself to progress monitoring or is a form of GOM. Do you an opinion on this? Can you guide me to research that I could read if you agree that it would not be a good idea?**

I do not see NWEA/MAP as competitive with AIMSweb. I know that some administrators see this as an “either-or decision.” I see them as complementary. As an assessment person, I believe schools need an independent, high quality achievement test, separate from their state standards. If this test allows for identification of *group* areas of academic strength and weakness, all the better for program evaluation purposes. To me, that is one of the strengths of NWEA/MAP. But again, it is not as useful as CBM for *frequent* progress monitoring and authentic screening. Frequent progress monitoring is crucial for high quality Tier 2 and 3 progress monitoring, for monitoring progress toward IEPs, and if progress monitoring is to be part of specific learning disabilities (SLD) eligibility in Response to Intervention (RtI). NWEA is both too infrequent for these purposes and also has not been validated to do so.

- **What type of assessment is represented by "Running Records" in reading?**
- **Talk to us about Running Records.**
- **I am still confused as to why a running record is not considered a tool for progress monitoring. Could you clarify? Why can't I use them as such?**

Within our field, there are generally four held purposes of assessment,

1. Screening/identification,
2. Diagnosis/instructional planning,
3. Progress monitoring, and
4. Program evaluation/accountability.

In my professional opinion, Running Records (RRs) are best used for “diagnosis or instructional planning.” They provide teachers information about what they do (or don't do) successfully as part of the reading process. This is *qualitative* information that I believe is important. However useful they may be for instructional planning, RRs have *not been validated* for screening nor progress monitoring, and are insufficient for program evaluation and accountability over. This is a question for research evidence, beginning with the most basic of technical questions, the reliability of the results. Given that most RRs lack sufficient standardization and are quite subjective, it would be difficult to determine whether the results are accurate. Additionally, I do think instructional diagnosis should be (a) *limited to those students who are not progressing within the curriculum*, and (b) should be conducted *with the diagnostic tests within the curriculum*. With respect to (a), it is easy to lose a lot of time testing students who are doing fine when in all likelihood, nothing will change instructionally as a result. If the intervention program or curriculum, does not have diagnostic tests, then I would assess using the strategies in Curriculum-Based Evaluation (CBE).

With respect specifically to RRs, I believe they are not time and cost efficient for *routine* use with all students. In general, I believe instructional planning should be done within the curriculum/program. Strong evidence-based interventions have placement and diagnostic tests. Finally, I don't believe all students need "diagnosis," especially when repeated over and

- **How do you address teacher feedback that CBM information does not align with diagnostic reading assessments such as Fountas & Pinnell Assessments?**
- **I work with many teachers/interventionists who bring a student's reading level from running records (Fountas & Pinnell, Reading A to Z, etc.) as evidence of a student's progress. Are you aware of any research on the validity of using these tools for measuring progress?**

Without trying to be flip, I would suggest that their feedback (opinion) is not accurate. This can happen for lots of reasons such as when a single piece of data doesn't not explain a few students performance. That is why people conduct research, to see what our eyes may not. A nice study conducted by Dr. Ben Ditkowsky with more than 500 students in Grades 1 and 2 found very high correlations between CBM and Fountas-Pinnell reading placements (.83 and .82 respectively). You can find his results on his website www.measuredeffects.com. Regarding the validity of F-P tests as progress monitoring tools, they have not been reviewed for this purpose by the National RtI Center (www.rti4success.org). Of course, my bias is that we use *validated* tools for their intended purposes.

- **What other GOM than AIMSweb for math do you recommend? Can you give me couple choices please?**

This is a difficult question to answer without more information. If you are looking to monitor progress, I would argue that the CBM mathematics measures are the methods of choice. If you are making screening decisions through Grade 8 and want authentic assessments, I would say the same tests are appropriate. If you are trying to do instructional planning, I would recommend the diagnostic and placement tests that come with the instructional programs.

- **Where do SLO measurements fit with MM and PM?**

I'm inferring that SLO in this instance means Standard Learner Outcomes. If this is correct, it is a Tier 1 method of assessing outcomes in content area courses in middle and high schools. If you read what CBM does, it focuses on basic skills within a General Outcome Measurement approach. Beyond the basic skills, I do not know of validated GOMs, and in particular, for secondary content area courses. By default, then, we are constrained to high quality MM approaches and must address all the assumptions of this approach. Of course, the first assumption of MM is that a validated instructional sequence drives progress monitoring test development. The second assumption is that reliable tests are constructed for each Standard. The list of requirements for high quality MM goes on and on and that's why MM presents logistics questions. Until validated content area GOMs are identified, if that is even possible, MM approaches are our best hope. But it is not at all easy.

- **Can you please say more about WE-CBM? You mentioned that it's a good measure for typically achieving students through grades 5-6.**

I believe that WE-CBM, having students write short essays in response to a story starter or topic sentence is a valuable screening and progress-monitoring tool to gauge basic writing skills, the ability to communicate an idea(s) in writing. We know that WE-CBM is less reliable than the other CBM instruments, in large part because writing is so influenced by interest and motivation. I recommend high stakes decisions be based on the at least three measures. It is easy to administer and score for Total Words Written, a good measure for most students, presuming we don't share the scoring method with students. For some students who produce, but not in a mechanically, semantically, or syntactically correct way, I recommend Correct Writing Sequences (CWS). I do *not* recommend scoring every writing sample using CWS. It works well for screening and progress monitoring through at least Grade 5. The measure loses its ability to identify high quality writing with older students, but for screening, that's not the target. Like many basic skills tests, one does not have to use a Benchmark approach, which combines screening and progress monitoring, for all students at every grade. With older students, I may use WE-CBM solely as a once a year screener or for individual student screening rather than universal screening. WE-CBM remains useful for progress monitoring when students are significantly discrepant.

- **Can you speak to the differences between Aimsweb Norms and Hasbrouk & Tindal's norms?**

The Hasbrouck and Tindal norms were accumulated "norms of convenience." The authors solicited all publishers of CBM reading probes and asked them to contribute their data to their normative base. *AIMSweb* contributed *all* of their extant normative data to the

Hasbrouck and Tindal norms, which made their data the largest single block of scores. Up until 2012, all CBM norms had been norms of convenience, reflected those customers who subscribed to data systems. Although AIMSweb users were generally reflective of the US population characteristics, it became the first company to use traditional methods to ensure a representative national sample. Whether the Hasbrouck and Tindal norms were largely AIMSweb data and regardless of whether the “old” AIMSweb norms were reasonable representations of the US, I personally had concerns about averaging information from different reading passages and if a national norm perspective is desired, see the recent normative efforts a major step forward.

- **We have struggled with assessing math progress with M-CAP, which doesn't seem as sensitive to change. Is it really useful to administer it on a weekly basis?**
- **We are having issues with using M-CAP to monitor progress. Even when writing ambitious goals, the rate of improvement is so small, the aim line is nearly flat, which doesn't lend itself to clear decision making during data reviews when determining if students are making adequate progress or a change in intervention is needed. How would you remedy this?**

Mathematics progress monitoring science lags behind what we know about reading progress, in large part because of the lack of quality mathematics progress monitoring probes. Sensitivity is a combination of at least two factors, the sensitivity of the scale and the power of the intervention. A good way to think about this is a bathroom scale. If the bathroom scale doesn't change, is it the scale or an ineffective weight loss program that is responsible? My first reaction, especially in mathematics is to examine the intensity of the intervention. When I see AIMSweb normative growth rates in mathematics, derived from Benchmark data, I see the potential to suggest lack of sensitivity. When I see progress monitoring graphs from students delivered high quality and intense mathematics interventions, I see dramatically different—and positive—rates of improvement.

Let's write our goals to reduce the gap. That will empower us to examine our outcomes and perhaps select and use more powerful interventions. It is easier to answer the question(s) about sensitivity when students are delivered powerful mathematics interventions with fidelity of implementation.

QUESTIONS SPECIFICALLY ABOUT CBM READING, MAZE, AND READING COMPREHENSION

- **What are other GOM PM measures for secondary students besides the MAZE? Students can appear ok on the MAZE when they learn how to figure out context clues, but still have a lot of struggles in the classroom with comprehension.**
- **Our district has been very dissatisfied with the MAZE for measuring reading comprehension. Can you speak to the validity of MAZE in measuring reading comprehension?**
- **What is the best response when people claim RCBM does not measure reading comprehension?**
- **Can you recommend a GOM that will measure a child's progress in the area of reading comprehension? As you know, we often need to intervene for students who do not have fluency issues but cannot comprehend what they've read.**
- **Is there a measure besides MAZE to assess comprehension?**
- **What is the best way to measure reading comprehension?**
- **How do we measure progress in the area of comprehension?**
- **What is a good GOM for reading comprehension?**
- **Some people do not think the Maze assessment is valid measure for comprehension. Thoughts?**

Clearly the sheer numbers of questions and comments around this topic echoes my own experience for 30 years. Educators' confidence in CBM is reflective of this single important question. It is why the fourth webinar and a white paper is directed to empirical as well as practical answers. A good answer, in brief, given the upcoming event, has multiple parts. First, remember that CBM is intended to assess **basic skills**. Basic skills are *necessary, but not sufficient* for content area success, including comprehension. In my opinion, there are no validated GOM tests beyond basic skills and, in particular, comprehension. Second, remember that Maze is a *silent reading test* that assesses *general reading ability* and is not (solely) a measure of comprehension. Like R-CBM, it is highly related to comprehension, but certainly not perfectly. Comprehension is affected by a number of variables besides reading skills, including motivation and interest, attention, language skills, especially vocabulary and familiarity with cultural idioms, and prior knowledge and life experience. Webinar 4 will address the issue of comprehension, and these sources of influence, which make comprehension much more difficult to measure, especially in an efficient way. Additionally, I would suggest most comprehension issues are Tier 1 issues, that we fail to do an adequate job to support understanding for *all students*, and that comprehension is highly dependent on what is being read and how it is being judged. Again, more in Webinar 4 for which you also will receive a white paper I've written entitled “the problem of comprehension.”

- **What are the benefits of maze tests? The negatives?**

Maze has the value of enabling very low cost universal screening for basic reading skills. If I have a sizeable proportion of older students (e.g., Grade 6 or above), most of whom are good readers, I may save considerable time by Benchmarking at least 1-year with Maze for purposes of universal screening in a multiple gating approach and general progress monitoring for all students. I would follow up at risk students identified by Maze with R-CBM to ensure motivation and effort. I also like Maze as a low cost corroborative progress monitoring tool for middle and high school students at Tiers 2 and 3. Once per month, I can administer Maze. If students are improving in their basic reading skills, I should see growth on both measures.

- **There is a lot of confusion within our building about using R-CBM as a tool to screen or as a tool to measure fluency. Some teachers are teaching children to read very fast to improve on the assessment. They say it is unfair to measure speed and not teach that specifically. I feel that if we teach effective reading skills, the results will be faster, more accurate reading. How do we get teachers to take the sole focus off of fluency and see the general value?**

This is a critical point that needs clarification and immediate attention. When students read aloud for a short period of time under standard conditions, R-CBM, we are assessing general reading ability, not “fluency.” A fluent French speaker is not a “fast” French speaker. Interpreting R-CBM as a “fluency”/speed test has no advantages and many disadvantages as you note. Your strongest teachers will push back—and appropriately—that good reading is not fast reading. The goal of good reading is *not* to read fast, but to read well. If treated as fluency, your interventions will likely lead to repeated reading. Practice is good, but repeated reading when students read poorly is not an especially powerful intervention. In sum, we are actually looking for an “authentic”

Chard, D.J., Ketterlin-Geller, L.R., Baker, S.K., Doabler, C., & Apichatabutra, C. (2009). Repeating reading interventions for students with learning disabilities: Status of the evidence. *Exceptional Children*, 75, 263-281.

Wexler, J., Vaughn, S., Roberts, G., & Denton, C.A. (2010). The efficacy of repeated and wide reading practices on students with severe reading disabilities. *Learning Disabilities Research & Practice*, 25, 2-10.

•

QUESTIONS ABOUT GOAL SETTING

- **Is it acceptable to use lower grade level R-CBM passages as progress monitoring for students who are far behind their actual grade level? If so, what is the best method for determining the appropriate grade level passages to use?**
- **What is the best way to set a goal using AIMSweb? Somewhere I have heard that 2 x the typical ROI is a way to set a reliable and attainable goal with intervention. Is that still recommended or is there a better way?**
- **Are there specific resources to help guide rate of growth for Special Ed students?**
- **How do you suggest that progress-monitoring goals be set for students in tier 2 or tier 3 who are performing many years below their grade level. Should there be standard adoptions of goal setting procedures set within a school setting to maintain consistency and to strive for ambitious growth?**

The next webinar on January 10th, 2013 will be an overview of goal setting practices for students in the “five families of progress monitoring:

1. Tier 1,
2. Tier 2,
3. Tier 3,
4. Special Education Students and IEP Goals, and
5. RtI SLD Entitlement.

The overarching idea in frequent progress monitoring is that our interventions (Tiers 2, 3, SE) need to *reduce the gap*. When we decide that the severity of the discrepancy is so severe we need to consider more individualized versus standard goals, a good place to begin for writing individualized goals is with the *AIMSweb* manual I wrote on goal setting practices.

Shinn, M.R. (2003). *AIMSweb Training Workbook Progress Monitoring Strategies for Writing Individualized Goals in General Curriculum and More Frequent Formative Evaluation*. Eden Prairie, MN: Pearson, Inc.

For goal setting for students in Tiers 2 and 3, as well as special education, you may also benefit from reading:

- Fuchs, L.S., & Deno, S.L. (1982). *Developing goals and objectives for educational programs*. Washington: American Association of Colleges for Teacher Education.
- Fuchs, L.S., & Shinn, M.R. (1989). Writing CBM IEP Objectives. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children*. (pp. 132-154). NY: Guilford.
- Shinn, M.R. (2008). Best practices in Curriculum-Based Measurement and its use in a Problem-Solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 243-262). Bethesda, MD: National Association of School Psychologists.
- 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.

Regarding the specific questions I've grouped together for this section, for students with IEPs, yes, "off-level" assessment is not only acceptable, but also highly desirable. In brief, for individualized goals, we need to know or determine three things:

1. *Expected Level of Performance*: This is the student's current grade. For a Grade 5 student, this is fifth-grade material.
2. *Present Level of Performance (PLOP)* as required by law. This is determined by the Survey Level Assessment (SLA) and judge normatively; the level of reading passages where the student in question reads as well as other students.
3. *The Goal Material*: The level of reading skill we would expect in the future if we were successful in reducing the gap.

Progress monitoring occurs in the Goal Material, *not* the PLOP. This is a difficult first step for the novice progress-monitoring teacher.

The most difficult part of goal setting is the criterion for the CAP, the Criterion for Acceptable Performance, for which there are four choices, local or national norms in a norms-based approach, scores that predict performance on high stakes, state standards tests in a standards-based approach, and rate of improvement (ROI). Each choice has advantages and disadvantages that are described. My bias is to set the CAP using local norms and my least preferred approach is to use ROI. I've not heard of expecting students to achieve double the ROI. To me, I would prefer that teams and teachers *understand the broader concept of reducing the gap* and what that may mean for students than rely heavily on mathematical formulas.

- **How do we get teachers to move away from always 80%?**

This tradition of 80% goes back to the classic piece on Writing Behavior Objectives (Mager and Pipe) written in 1971 or 1972. The idea implicit in the criterion is that we want students to do "better" but they don't have to be perfect (e.g., 100%). This has common appeal, but is inherently ridiculous in practice. At the most basic level, this is an "accuracy" criterion. But accuracy is the lowest level of proficiency in a learning hierarchy. I might "accurately" solve an addition problem, but I want to do some things with "automaticity."

Here is an example of how a good idea can be "weird" in practice.

- Do you want children to be toilet trained with 80% accuracy? No.
- Do you want children to cross the street successfully 80% of the time? No.
- Likewise, we don't want students to reading with 80%. They will have great difficulty understanding and reading will not be fun and rewarding.
- Do we want students to solve mathematics facts problems with 80% accuracy? No. That is not a foundation for developing number sense and ensuring ease with procedural algorithms.

And accuracy is insufficient for maintenance and generalization. If a skill is important, in and of itself, and as a basis for future learning, we must do more than be accurate and only "most of the time." If we want students to be able to regroup fractions, they will have great difficulty if they are not only accurate 100% of the time, but they can use these algorithms with ease and efficiency.

$\frac{1}{4} = x/16$ needs to be automatic.

One of the principal reasons that CBM was developed was that when major funded special education laws were written in 1975 with regulations in 1977, the IEP was a great idea and there were concerns that the prevailing goal setting technology (e.g., short-term objectives with 80%) would be insufficient. Federal \$\$\$ were allocated to develop a scientifically based alternative—that alternative became CBM. For more history see:

- Deno, S. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure*, 36, 5-10.
- Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219-232.
- Deno, S.L. (2003). Developments in Curriculum-Based Measurement. *The Journal of Special Education*, 37, 184-192.
- Deno, S.L., Mirkin, P., & Wesson, C. (1984). How to write effective data-based IEPs. *TEACHING Exceptional Children*, 16, 99-104.
- Jenkins, J. R., Deno, S. L., & Mirkin, P. K. (1979). Measuring pupil progress toward the least restrictive environment. *Learning Disability Quarterly*, 2, 81-92.
- Jenkins, J.R., & Fuchs, L.S. (2012). Curriculum-Based Measurement: The paradigm, history, and legacy. In C. A. Espin, K. McMaster, S. Rose & M. Wayman (Eds.), *A measure of success: The influence of Curriculum-Based Measurement on education* (pp. 7-23). Minneapolis, MN: University of Minnesota Press.

How do we get teachers to move away from 80% criterion—and the larger issue—to get away from an emphasis on short-term objectives that are difficult to observe and measure? My first answer is not intended to be facetious. I would suggest that district leaders *require and train* teachers to write better goals, beginning with special education IEP goals. Most persons write their goals based on their (lack of) training and when given permission to write fewer, more functional goals *and* training, adoption is relatively straightforward. A *volunteer* approach to changing goal setting practices will invariably fail simply because change is difficult.

- **For a student who is below average, if reading progress is made at the rate of the typical student, is this adequate? Gap is not closing. What should then happen?**

To me, the goal of intervention (e.g., Tiers 2 or 3) is to reduce the gap. So, my answer would be if the goal is to reduce the gap (the aim line) and the actual rate of progress is not acceptable, we change something. I have a hierarchy to consider:

1. Assess whether the intervention is implemented with integrity.
 2. If yes, consider whether the student is appropriately placed in the intervention program and/or is not lacking some essential skill or strategy. If yes to the first point and no to the second point, then consider the need to change instructional placement or reteach the critical skills or strategies.
 3. If yes to 1 and 2, assess whether other students who are receiving the same intervention are benefiting. Sometimes an advertised intervention is not as effective as claimed, even when implemented with fidelity. If no, then the intervention needs to be changed for all the students.
 4. If yes to all 3 questions, consider how you might modify the current intervention. I've suggested using a variation of what is called the Instructional Planning Form (IPF), a simple way to describe the current intervention, as a basis for identifying potential variables that could be changed.
- **My experience that general education teachers accept progress monitoring more readily than intervention specialist. The reason....intervention specialist do not understand "how to change their instruction", what does this look like, how do I know that this works.....thus progress monitoring. But PM is scary to them....I have had them cry when the trend analysis shows "flat lining" or regression.**
 - **How do you interpret AIMSweb to adjust teaching?**

Changing intervention is often difficult, in large part, because most of us have not *had training in how to do so*. For some educators, especially with specialized training, this can be even more difficult because our specialization has often come at the cost of *adoption of an intense commitment* to a particular philosophy or instructional program. As one example, what if one's training suggested that students must be taught according to their learning style? Although this common perspective is without evidence for effectiveness, imagine that your learning style instruction has generated a "flat line." Might you have difficulty changing, first because of your commitment and training in learning styles, and relatedly, you lack of training in how to change instruction? Lynn Fuchs demonstrated almost 30 years ago that teachers who monitored progress became more open to consultation and less dogmatic about what worked. We can do this systematically by coaching and in particular, through use of the Instructional Planning Form (IPF), a no-cost, time efficient way to describe instructional modifications to support change.

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

- **We are currently using AIMSweb and are discussing what scores are best to use for progress monitoring of students with special needs. National norms, district norms, or cut scores. Many people are leaning towards cut scores but I am confused because some of us believe cut scores are for benchmarking. Can you clarify and advise, thank you.**

See the previous answer for some assistance. I would not use a Standards-Based Approach, the cut-scores you refer to in your question, as the *primary* method for determining the CAP. As I hope to communicate in the upcoming webinar, I believe the cut scores can provide some “value-added” information, but to me, these scores are best used for program evaluation purposes and *not even as primary criterion for screening*.

- **Our district uses spring 25% for the year-end goal. This seems low in my opinion. Shouldn't we have higher expectations for our students?**

This question is difficult to answer without more information. To me, this could be a suitable goal for either students at Tier 2 or Tier 3. Let's assume that Tier 3 intervention is provided to students below the 10th percentile. A goal of the 25th percentile by year's end would be outstanding. Attaining this goal may mean the student is an average reader and no longer requires specialized intervention. There may be the concern about maintaining growth, but I can remedy this by continuing frequent progress monitoring, which I strongly recommend with any student as part of their transition plan.

This also could be a suitable goal for students who may be targeted for Tier 2 intervention. I prefer to target students below average (i.e., < 25th percentile as a starting point) for Tier 2. If the specific student meets this goal, by definition they are an average student who no longer requires intervention. But to me, your goal for Tier 2 needs to be related to your screening criterion. What defines the problem should define when it is no longer a problem.

QUESTIONS ABOUT CBM AND SLD/SPECIAL EDUCATION ELIGIBILITY

- **Can Progress Monitoring be used to specifically identify learning disabilities to qualify students for special education services?**

Of course, the hallmark of RtI is the capacity to measure “*response*” or *progress*. Failure to monitor progress in the basic skills—and what better tools are useful for this purpose than CBM?—means you can't use RtI. The unique feature of RtI in SLD identification is the dual discrepancy, in educational need as assessed by the achievement performance discrepancy, and (lack of) educational benefit, as assessed by a low rate of improvement given appropriate instruction.

- **If you have students receiving research-based intervention and have been appropriately progress monitored and have made adequate rate of growth on their graphs throughout several years but will take many years to close the gap. When should we consider a case of learning disability be considered?**

Another very important, but complicated question. The general principle is that legally we are required to consider special education when we “suspect” the student has a disability. I don't want to give this question short shrift, but let me attempt a brief answer. First, our goals for Tiers 2 and 3 should be to reduce the gap. If adequate progress is the definition of our intervention goal efforts, we should not expect to reduce the gap and I would argue our interventions will not be sufficiently intensive. I will assume the student has the first of the dual discrepancies, severe educational need. I will assume that adequate progress in this instance means the student is reducing the gap, but not fast enough. My perspective is this. If the student is severely discrepant *and* I suspect that the student has not had appropriate instruction or I cannot rule out other “Determinant Factors” like English Language Learning, I should not suspect a disability. If after a period of appropriately intensive instruction the student is failing to reduce the gap *and* I can rule out the other determinant factors, I may suspect the student needs to be consider for special education eligibility. However, this decision also puts the burden of proof on the school to show how special education intervention would be different than the interventions currently being delivered, say, for example, through Tier 3.

QUESTIONS ABOUT EARLY LITERACY AND MATHEMATICS

- **For kindergarten students, would you recommend using AIMSweb LSF, IGDI's, or a combination of both as GOM's?**

AIMSweb is especially suited for time and cost efficient early literacy assessment. I have concerns that some conventional practices with multiple measures at each Benchmark has shown little improvements in decision making and has come at the cost of too much time.

When schools ask me, I suggest the following for K:

Fall Benchmark: Letter Names, and, if accessible, the IGDI Vocabulary Measure

Winter Benchmark: Shift to Letter Sounds. In most communities, I would do an additional Benchmark with LS in November or if I do not place entry kindergarten students into intervention based on Fall Benchmarking—not something I am recommending, but it happens quite often—I may delay my Benchmark until November and use LS.

Spring Benchmark: Letter Sounds *or* Nonsense Words *or* Highly Decodable Passage

I do not routinely recommend tests like PSF because I believe they are best used diagnostically *if* students are not making progress. There is nothing wrong in collecting PSF data, but in the interest of time, I would prefer to minimize the assessment of all students, many of whom are progressing.

- **There are 10 potential different CBM tests for 1st grade looking at math and reading. To keep this "short" and to keep teachers teaching what do you recommend 1st grade teachers to use.**

See the answer to my previous question. I strongly believe that TEL and TEN need to be time and cost efficient in terms of the number of measures administered and time spent in assessment. Saving minutes adds up. At Grade 1, consider highly decodable passages in Fall. Administer other measures only if students fail to perform adequately. For TEN, consider only Missing Number *or* Quantity Discrimination.

- **Why hasn't there been a more advanced phonics skill past Nonsense Word Fluency to be a general outcome measure to predict future reading skills. I find in my practice that many students are established on CVC words but are very low in reading fluency. Mastery Measurement could be used to teach phonics skills but it would be nice to also have another progress monitoring tool between an early phonics skills (CVC) and oral reading fluency**

I concur. My bias here is to employ authentic highly decodable passages. These passages allow both quantitative and qualitative judgments about reading skills to be made in an authentic way.

QUESTIONS ABOUT SECONDARY USES OF CBM

- **We are having trouble coming up with meaningful GOM measures at the 9-12 level. Do you have advice on these grade levels?**
- **Will you talk about using ORF (R-CBM) at the middle and high school level? I have heard that new research is suggesting that ORF isn't as effective at the upper grades versus the elementary grades.**
- **For the upper grades: using R-CBM as a screener only gives you an 8th grade passage so how do you know what the criterion/percentile is for 10th and 11th grade when students are now using reading to learn content vs. learning to read. How is this used as a general outcome measure in the high school setting?**

The current Grades K-8 CBM measures work just fine, but your question is quite common. I've been working on use of CBM at the secondary level in what has become MTSS since 1982. You can find some narrative around some of this in an article I wrote on Secondary RTI for the *RTI Answerbook*.

Shinn, M.R. (2008). RTI at the secondary level. In S. L. Fernley, S.D., Norlin, J. (Ed.), *What do I do when...The answer book on RTI*. Horsham, PA: LRP Publications.

CBM is designed to provide simple, sound, and seamless measures of *basic skills*, first and foremost for use in frequent progress monitoring (PM), especially at Tiers 2, 3 and for IEP goals. The measures are applicable for PM as long as students are discrepant in their reading skills. Leading to: WHAT IS DISCREPANT?

Key concept, first, is basic skills (BS). They are necessary, but not sufficient for content area success. One can have BS in reading, but not understand a 10th grade history text. This type of problem, to me, and probably to many others, is NOT a special education problem, and may not even be a Tier 2 problem. Lots of students have difficulty understanding a 10th grade history text. The solution is better core instruction (for everyone) in navigating expository/informational text, taught and supported by *all* content area teachers. We know how to do this, but sadly, most of our content area teachers are not taught these methods. So we have to do it in-service instead of pre-service.

CBM is designed to PM and SCREEN for students with basic skills needs. That's why the probes only go up to Grade 8, when, by most conventions, few schools or curricula have stopped teaching these basic skills.

Now, what is discrepant? K-8, for years, I've encouraged a *normative approach* to defining the discrepancy, and in fact, a *local normative approach*. Of course, it doesn't what which norms if the local or national norm is similar. It does matter if the local norm is significantly higher (or lower) with respect to how teachers, parents, and students experience and perceive a discrepancy.

Post Grade 8 (i.e., high school) I strongly encourage a *standards-based approach* to define the discrepancy. Standards-based requires schools to *define the minimum reading standard for proficiency*. As a point of first argument, I suggest considering whether holding students to an end-of-grade 7 standard would be satisfactory for the community. If all students could read at least as well as end-of-grade 7, could we then work on supporting students with the necessary supports they will need to ensure content area reading and understanding success? It may seem like a low bar, but in fact, it is reasonable, given our relatively low national average reading proficiency. Students below an end-of-grade 7 may need intensive intervention (and of course, we wouldn't screen if we don't intervene). To me, this is foundational.

DON'T SCREEN IF YOU DON'T INTERVENE!

To identify students--and to determine my goal for students who receive intervention--I would individually screen targeted students against an end-of-grade 7 norm on Grade 7 passages. My goal for intervention, presuming the student's skills are not so low as to require more individualized goals, would be to have the student read as well as an end-of-grade 7 student.

My theme: Discrepant from Standards-----**TREAT** with Intensive Intervention

- **What suggestions do you have for secondary schools trying to use GOM effectively with the entire population and with delayed learners?**

I'd say this is not a GOM question but an instructional question. In high risk communities, it is imperative that we strengthen the core language arts curriculum under the circumstances you describe. My bias would be to use an evidence-based program with a double period or double block every day with well-trained teachers and really powerful core programs. We know how to do this and it has been done.

- **What do you recommend to use in the middle school progress monitoring, R-CBM or MAZE?**

I answered this question partially in a separate question about Maze. I would prefer R-CBM to the greatest degree possible because I can judge student effort. However, I also find Maze a time and cost efficient regular (monthly) corroborative progress monitoring tool for struggling middle and high school students with severe basic skill deficits.

QUESTIONS ABOUT EARLY CHILDHOOD AND PRESCHOOL USES OF CBM

- **What is the best way to measure vocabulary growth in young children (Kindergarten)?**

I remain convinced that a receptive vocabulary measure with lots of items on key words (e.g., every day words) remains the best choice. This is best operationalized by the Vocabulary measure from the Individual Growth and Development Indicators (IGDIs).

- **Do you have any suggestions for progress-monitoring for EC? We have piloted IGDIs, but teachers did not find them useful and we did not have norms because we have an all special needs school.**

To date, I find the Vocabulary measure the most useful IGDI for preschool. When you say useful, again, it depends on assessment purpose. IGDIs are *not* intended to be useful for diagnosis/instructional planning and that it was teachers often react to: they want

(legitimately) to know what to teach. For screening purposes, with a special needs school, you don't need to do this. Students are already identified. If you want to measure progress attributable to your language intervention program, and with a special needs school I'm presuming you have a powerful one, then you can monitor progress with the Vocabulary IGDI.

- **Would it make more sense for us to create MM for specific skills we want them to have?**

I would say no. I have problems when teachers create tests. It takes time and training, of which many of us have little. We don't have enough validated general outcome measures (GOMs) but that doesn't mean we shouldn't do our best to monitor progress. My strong suggestion is to monitor progress with the tests that come along with the intervention programs. The evidence-based ones have (imperfect) ones. If you are not using evidence-based interventions, especially language, in a special needs preschool, a problem may be other than a progress monitoring test. It is asking a lot of teachers to develop their own curriculum as well as their own tests.

QUESTIONS ABOUT CBM AND COMMON CORE STATE STANDARDS

- **Formative Instructional Practices and Formative Assessment from deconstructing Common Core ELA and Mathematics Standards, many school districts are training teachers to set learning targets, collect evidence of students learning, encouraging students to take ownership of their learning and the formative assessment is differentiated based on student need, performance based and thus is mastery monitoring. How do we get districts to understand the correlation between GOMs and formative instructional practices? I hope that this question makes sense. Ohio is moving toward performance criteria assessment. I believe that the real question is "what is good quality instruction". What performance criteria does a teacher use to evaluate her instruction and the student can evaluate the instruction.**

I can address the part about formative assessment. The first webinar in the series I completed was on the relation of CBM to the Common Core State Standards (CCSS). A white paper I wrote also was released as part of that presentation. My conclusion was that CBM is both consistent (i.e., content valid) with many of the CCSS anchor and foundational standards, especially for K-5, and also is complementary (consequential validity) to decision making to attain the CCSS. The assessments that will be aligned to CCSS are not screening tests nor are they progress monitoring tests. They are accountability and program evaluation tests. Schools still need to make important decisions within an RtI/MSTSS framework and primary among them is the ability to do frequent progress monitoring. In the white paper, I expressed concern that important gains in simple, scientific, and practical assessment of basic skills will be discarded in the quest to attain the CCSS. Of course, outside of the basic skill areas where CBM works so well, the challenges that quality Mastery Monitoring (MM) presents must be addressed. MM is not wrong. It is just not all we need to do and in areas like content area learning, it is our only viable technology.