NWEA Research Brief

From General Outcome Measurement to Mastery Measurement: A Case for Instructionally Useful Information

Cindy Jiban, PhD

John Wood

December 2016

Northwest Evaluation Association

Partnering to help all kids learn

Contents

General Introduction 1

Executive Summary 1

Research on General Outcome Measures (GOM) 2

Moving Beyond GOM to Mastery Measurement (MM) 5

Curriculum Sampling: A GOM with MM Characteristics 6

Traditional View of Mastery Measurement: Low Rigor, Instructional Sensitivity 8

A New Vision for Mastery Measurement: Rigor and Structure 9

Conclusion 10

References 12

General Introduction

The growing popularity of personalized learning fostered by the wider availability of quality digital curriculum has increased the need for screening and progress monitoring for all students. Response to Intervention (RTI) is good instructional practice and, as such, has application beyond the narrow audience of students who are deemed to be at risk. This paper looks at issues around progress monitoring in the narrower context of traditional RTI. Most of the progress monitoring research has focused at oral reading fluency, the most prevalent progress monitoring tool. The paper discusses the research that shows that oral reading fluency and other General Outcome Measures (GOM) cannot provide valid and reliable information for high-stakes decision making over short durations of time.

As an alternative, the paper discusses Mastery Measurement (MM), which is more instructionally sensitive, provides more information for formative decisions, and applies across a wider range of curricula beyond the early grades. Finally, the paper discusses a model for MM that meets one of the recommendations Edward Shapiro made in 2013 in a special issue of the Journal of School Psychology, which focused on the problems of progress monitoring (Shapiro, 2013, 64-65). The recommended solution is not only applicable in the narrow context of RTI application, but in the broader application of RTI practice to all students across grades.

Executive Summary

The decisions RTI teams make based on screening data, progress monitoring data, and other data sources are relatively high-stakes decisions: recommendations for special education, recommendations for tier change (intensity of intervention), and recommendations for changing the intervention program. Additionally, there are other lower-stakes decisions that educators can make based on progress monitoring data. These decisions are more formative in nature, including what to teach next and when to move on.

Educators have primarily relied on GOM, which often serve the dual purpose of screening and progress monitoring. While GOMs correlate well to overall competence in a domain as measured by other instruments, they provide limited formative information, particularly when the indicator is not the instructional focus. For example, oral reading fluency correlates strongly with reading comprehension while fluency is increasing and can be a proxy for overall reading ability, including comprehension. However, it cannot provide instructional information for teaching reading comprehension.

Some GOMs use a curriculum sampling content design in which a year's curriculum is represented on alternate test forms that sample from all grade level skills. These GOMs estimate progress toward attaining grade level proficiency across the year by considering the overall percentage of correct answers on each form. They may provide some formative information depending on the degree of overlap between what is currently being taught and what is sampled from the curriculum.

All GOMs become less useful for formative decision making as curricula become broader and more complex in higher grades. MMs, on the other hand, focus specifically on the skills that are being instructed at any one time, and provide information about what to teach and when to move on. While MMs with their granular information are difficult to use to get a global view of progress within a domain, they may be equally useful as GOMs in providing information to complement good interim assessments when making high stakes decisions about treatment and placement.

Progress monitoring is characterized as repeated measurement, but research indicates that more frequent administration of GOMs does not provide better information for making important decisions. Rather, it is administration of a high-quality GOM over longer time periods that improves decision making. With research converging on durations of at least eight—and, more reliably, 12 to 14—weeks, the information for decision making can just as effectively come from interim assessments.

Given that progress monitoring need not bear the primary weight of estimating general growth, it makes sense to complement a rigorous interim assessment with an instructionally sensitive MM to provide educators both growth information for high-stakes decisions and instructionally useful formative information.

Using MAP®, which can be administered every 12 to 15 weeks, and Skills Navigator® together provides this combination of reliable growth data and instructional information. They are a powerful solution for progress monitoring.

Research on General Outcome Measures

The essential data-based decision making components of an RTI or a Multi-Tiered System of Support (MTSS) model include: 1) universal screening and benchmarking, often thrice yearly, and 2) progress monitoring. Progress monitoring has typically meant weekly or bi-weekly administration of GOMs, intended as a method of determining a student's growth or responsiveness to instruction over short intervals (e.g. six weeks). However, current research is converging on the reality that growth estimates from GOM progress monitoring are technically indefensible for individual decision making for intervals shorter than 10 to 12 weeks. This emerging understanding makes GOM data redundant to that offered by a screening and benchmarking tool, and raises the question of what additional information we should expect from a progress monitoring measure.

GOM progress monitoring has been a centerpiece of RTI. Its history reaches back before RTI, largely under the name of Curriculum-Based Measurement (CBM), to an initial purpose of helping special education teachers make data-based instructional adjustments as they monitored progress toward Individualized Education Program (IEP) goals (Deno and Mirkin 1977). These were simple, brief measures designed to be sensitive to student growth, with scores charted across time (Deno 1985). GOMs, especially CBMs, have long contrasted themselves against skills MM. While a GOM uses the same measure repeatedly across a school year and focuses on the slope of growth charted, an MM uses a measure only until a skill is mastered, before replacing that measure with one of a next skill. The RTI paradigm was largely developed from a premise of GOM use, although use of MM for progress monitoring has been a recognized alternative via training and tool review agents, such as the National Center on RTI.

Research literature accumulated around CBM, prompting Fuchs (2004) to frame up a categorization schema for broad CBM types and for stages of research necessary to support their use in making various decisions about students. With regards to types, Fuchs noted that some CBMs are "curriculum sampled", while others are what she termed "robust indicators," chosen not for their representative sampling from skills taught, but for their strong and useful correlations to broader measures of proficiency. No CBMs are MMs; however, they might sample skills from the curriculum, but they include this same sample in all alternate forms or "probes" across the whole year. The data from any CBM, whether curriculum-sampled or not, is a GOM: total correct is the data point, not scores on subskills.

Fuchs also laid out in her schema stages of research relevant to CBM. Stage one is about their use as point-in-time measures: reliability of alternate forms, criterion validity, even classification accuracy. Stage two research is about technical features necessary for estimating student growth, as in continuous progress monitoring: reliability, precision, and validity of slopes of growth generated from time-series data. Fuchs called for more focus on later stages of research, noting insufficiencies beyond stage one.

Much evidence exists from stage one research focusing on criterion validity: do scores on this short measure correlate strongly to other valued measures of the larger construct? For oral reading, correlations between GOMs and broader measures of reading proficiency, including comprehension, have been shown repeatedly to be high for readers in primary grades and at lower levels of proficiency (Wayman et al. 2007). However, GOMs in math typically show more moderate criterion validity (Foegen, Jiban and Deno 2007; Ketterlin-Geller, Gifford, and Perry 2015; Sisco-Taylor, Fung, and Swanson 2015). Generally, GOMs in math have been of the "curriculum sampled" rather than the "indicator" type, and researchers have long discussed the likely inevitability of this, at least after the youngest years where basic numeracy develops (Foegen, Jiban, and Deno 2007). Math is much more a collection of overlapping topics than one coherent progression; assessments that align to scope and sequence of instructional topics work better.

After Fuchs' call for more stage two research, a set of studies appeared challenging the technical adequacy of short term CBM slopes—particularly the oral reading fluency measure most successful at stage one—for capturing an individual's growth reliably over time (Hintze and Christ 2004; Christ 2006; Jenkins, Graff, and Miglioretti 2006). They noted that students' correct words per minute bounced up and down across the time-series data graph, making a line of slope drawn through those data very imprecise. Recall that an individual's slope of growth is the summary data piece that rolls up to decisions about whether or not the student is responding adequately to instruction or intervention; decisions about changing tiers or intensity often rest primarily on this slope. But research began to question: how many GOM progress-monitoring data points are needed for this slope to be reliable and precise enough? How soon before a school could reasonably make those short-cycle data-based adjustments to a student's instruction, the hallmark of the RTI model? Two moving parts to this question were parsed out: the administration schedule, or how long and how frequently to progress monitor, and the quality of instrumentation, such as passage sets.

GOM progress monitoring offers no evidence that it can shorten the cycle of data-based decision making over a technically sound, thrice-yearly screening tool designed for growth measurement.

While the original premise of GOM progress monitoring was that it would allow educators to make informed adjustments after a few weeks to a month, this has not borne out. By 2006, nine or ten weeks of instruction was asserted as a minimum for determining an individual's slope with enough precision that it could be reliably compared with a target or aim slope of growth (Christ, 2006; Fuchs et al. 2006, as cited in Jenkins, Hudson, and Lee 2007). The administrative frequency associated with these assertions of nine or ten weeks was at least twice a week. However, Ardoin and Christ (2009) found that the error associated with individual slopes, even at twice a week for ten weeks, was so high that a slope that looks like an increase of 1.5 words a week can only reasonably be interpreted as somewhere between 0.28 and 2.44 words per week. In other words, "the student's response to instruction was somewhere between inadequate…to excellent" (Ardoin and Christ 2009, 280). In another study, Christ et al. (2012) showed that 14 weeks of progress monitoring would be necessary with a once a week frequency. By 2012, numerous studies had begun to focus on the central importance of duration of progress monitoring: what matters most is how many weeks of instruction are covered, not how many data points per week are collected. This led to investigations of intermittent (e.g. monthly) administration (Jenkins, Graff, and Miglioretti 2009; Jenkins and Terjeson 2011) and even a pre- and post- design, where no progress measures were administered in between, since the progress-monitoring process did not contribute additional information to the estimate of a student's growth (Christ et al. 2012). The general finding was that 10 to 14 weeks of instruction needs to occur between the initial to final administrations of a cluster of passages, and that continuous weekly administration does not shorten this number of weeks.

These findings about administration schedule have something of a co-dependent relationship with the quality of the particular instruments used. Quality of instruments for more continuous progress monitoring is strongly affected by their passage or "probe" equivalence, including not just alternate form reliability, but equivalence of scores generated. Ardoin and Christ have argued repeatedly that wide-spread misunderstanding of this issue has led to broad adoption of what are marketed as "reliable" measures that actually fail to minimize the "bounce" in a student's scores over time (Ardoin et al. 2013). Specifically, they found that oral reading passage sets that were leveled into sets based on readability formulae, such as DIBELs (Dynamic Indicators of Basic Early Literacy Skills), led to poorer quality data sets than those passage sets developed more empirically from students' readings (Ardoin and Christ 2009). For poorer data sets, the minimally viable duration needed in the data cycle increases to even more than 10 weeks.

Ten or more weeks begins to approximate the cycle of screening or benchmarking. It is critical, given the budget and test-time pressures that schools currently face, to make explicit this key finding of research: *GOM progress monitoring offers no evidence that it can shorten the cycle of data-based decision making over a technically sound, thrice-yearly screening tool designed for growth measurement.*

The general finding was that 10 to 14 weeks of instruction needs to occur between the initial to final administrations of a cluster of passages, and that continuous weekly administration does not shorten this number of weeks.

On methodological rather than empirical grounds, Paris (2005) has also made the case that the correlation between oral reading fluency and reading comprehension has been oversold, making this measure less ideal than some have claimed for monitoring progress on a general outcome of interest. Paris' argument centers on a discussion of constrained versus unconstrained skills in reading development. Letter naming, for instance, is a highly constrained skill, with a period of growth sandwiched between an obvious floor (knowing no letters) and a clear ceiling (knowing 26 out of 26 letters). The efforts to find robust GOM indicators within the RTI progress monitoring world has relied on finding strong correlations, and Paris argues that the strength of many of these correlations is fleeting at best: these measures only correlate strongly during the short period between total non-mastery and complete mastery, after which they offer no meaningful variance. Treating constrained skills as unconstrained—for instance, using parametric techniques with data that is not normally distributed—has led to results that just don't hold up under scrutiny.

Oral reading fluency is a somewhat constrained skill. Paris argues: it is constrained by "speed of speech production", as well as automaticity with word decoding. More importantly, its correlation to the broader construct of reading comprehension is based on an "asymmetrical codependency"; for novice readers, "lack of oral reading fluency is correlated with…lack of comprehension, but fluency is not necessarily correlated with comprehension among skilled readers" (2005, 192). While it is important to become sufficiently fluent to comprehend text, it is not important to read faster still. Instead of working as an ongoing indicator of growth in overall reading, a slope that keeps working as a gauge of improvement, fluency might be better characterized as a prerequisite skill that can be "mastered" to sufficiency for a certain level of text.

Paris' notion of constrained skills is a lens to view all general outcome measures. They are useful and informative only during the period of time that the skill is being acquired. Steady improvement, at a sufficient rate modeled by slope, may not even be linear as GOM theory posits. As skills are acquired, the rate of improvement may slow as learning approaches the ceiling that is represented by solid acquisition of the skill, even though that skill is not completely constrained. Consider reading comprehension as an example. Adult readers may become more and more skilled as readers over time, (e.g. a student at the end of an undergraduate program and the same student at the end of a graduate program). However, this sort of growth is much more nuanced and subtle than growth in comprehension at earlier ages. The MAP test of reading comprehension also suggests that growth in somewhat constrained skills is more asymptotic than linear.

Moving Beyond General Outcome Measurement to Mastery Measurement

If progress monitoring data based on "robust indicator" GOMs do not offer a defensible shorter cycle of data on a student's growth, and if they are GOMs that do not parse out which skills are weak, then what *is* the logic of progress monitoring? In a 2013 special issue of the Journal of School Psychology, a devastating set of issues around the use of CBM progress monitoring was recapped (Ardoin et al. 2013; Christ et al. 2013). Shapiro (2013, 64-65) took stock of the important issues facing RTI implementation at this juncture. He offered three recommendations. The first two are quoted here and the third calls for more studies of whether progress monitoring data adds value to decisions based on screening data.

1. *Increase efforts to demonstrate the fidelity of implementation of intervention and rely more on universal screening than progress monitoring outcomes to determine impact of interventions.*
2. *Provide a progress monitoring measurement system that includes assessment instruments that yield scores that are sensitive to instructional change, meet psychometric standards for reliability and validity, and are predictive of universal screening and other important criterion measures of intervention outcome.*

On a positive note, both Shapiro (2013) and Ardoin et al. (2013) noted that various screening and benchmarking tools were technically well supported for the purpose of growth estimation—including both CBM-type screeners and content-aligned CAT-type screeners. This might free up the progress monitoring measures from trying to bear the primary weight of estimating a student's general growth in reading or math, Shapiro notes. In fact, Shapiro et al. (2012) studied decision making over a two-year period and found that 80-85% of decisions were consistent with the single data point of the universal screener.

In general, the desired outcome for student performance is, at minimum, proficiency on grade level standards. Thus, it is critical for performance on a "robust indicator" or other CBMs to correlate well with performance on external measures, like state summative assessments. Put another way, educators often use GOMs to predict performance on external measures, such as state summative assessments. In an era of high-stakes testing, educators have come to rely on this predictive information to guide instruction. However, the information provided by progress monitoring—adequate or inadequate rate of change (slope)—does not really add much predictive power to the information provided by measures of student status (Silberglitt, Parker, and Muyskens 2016). Rather, it is student discrepancy in terms of performance level that predicts success. Given that, why include slope in a model if it adds no predictive value? Silberglitt, Parker, and Muyskens (2016) concluded that though it adds no predictive power, slope does add information that contributes to decision making based on the Shapiro et al. (2012) finding that progress monitoring data influenced decisions about 15-20% of the time.

Since educators can't get reliable short cycle information about growth or progress to inform decisions, and since progress monitoring information does not add predictive power for student success, it is logical to focus on choosing progress monitoring tools that provide formative information. Curriculum sampling and MM may be more useful. Though GOMs based on curriculum sampling are primarily designed to estimate progress toward attaining grade level content across the year, they may provide formative information. MM focuses specifically on the skills that are being instructed at any one time and provide information about what to teach and when to move on.

Curriculum Sampling: A GOM with MM Characteristics

The Iris Center at Vanderbilt University provides an overview and comparison of MM and curriculum sampling as progress monitoring tools. The site points out the strengths of curriculum sampling GOMs and MMs as:

1) the skills are aligned with the curriculum

2) skills may be assessed frequently

3) the assessment results can be used in planning

Curriculum sampling takes the set of skills that comprise the year's curriculum and samples those skills in repeated testing across the year. Progress is charted by graphing the increase in percent correct from early in the year probes to those later. Curriculum sampling has primarily been used in math since there are not good "robust indicators" available beyond the earliest levels. Sometimes, the whole curriculum is not sampled but a strand of the curriculum like computation in math. Most studies of curriculum sampling in math have focused on computation rather than the entire curriculum (Tindal 2013).

Mathematics may be better viewed as a set of constrained skills, as well. This application of Paris' frame echoes a parallel thought proposed by both Deno and Fuchs, pioneers of the CBM paradigm: mathematics may not lend itself to measurement by a unitary, continuous indicator of overall proficiency—there may be no corollary to oral reading in mathematics (Foegen, Jiban, and Deno 2007; Fuchs 2004). Instead, they propose measuring repeatedly a collection of skills sampled from the curriculum. However, for a GOM, a student's performance on this sample rolls up to just one overall score as data point, and growth is charted as if the construct measured were in fact unconstrained. Paris' critique applies: a fixed set of constrained skills is not the same as an unconstrained skill, and treating them as such will mean that correlations drop significantly for more proficient students. As groups of kids successfully master individual skills—reaching their ceiling—fewer and fewer skills bear the remaining variance among student scores. Differences on a small set of particular skills get represented as differences on the broader construct of overall math proficiency. Viewed through the lens of Paris' important critique about constraint, this means that scores from a general outcome measure correlate to the outcome less and less generally, as students grow.

Though curriculum sampling has not been investigated to the level of "robust indicators," many of the issues around decision making are the same: variability of scores, sensitivity to instruction, and dependence on equivalence of alternative forms. RtI decision making faces the same caveats that limit the use of "robust indicators". Curriculum sampling also faces the issue of finding a balance between coverage of the domain and the desire to keep frequently administered probes short. This balancing becomes more and more difficult as the curriculum becomes broader in higher grades and is less focused on developing core competencies, like reading fluency and computation fluency. In higher grades broad skills which develop over time are less prevalent and replaced by knowledge acquisition including domain specific vocabulary and concept development. This knowledge acquisition may include discrete math algorithms that as Paris noted are constrained. Silberglitt, Parker, and Muyskens (2016, 274) notes "… that GOM assessments tend to be more applicable with younger students or acquiring skills, whereas SMM [sub-skill mastery measures] assessments tend to be more applicable with older students or the acquisition of knowledge." This transition begins in the elementary grades.

In terms of making formative instructional planning decisions, curriculum sampling has some drawbacks. First, to provide good information about all aspects of the curriculum, the sampling probes would have to cover the breadth of the curriculum with sufficient depth to make inferences about when to move on. However, creating such probes is not the purpose of curriculum sampling. The central purpose is to deliver an overall score stable enough to graph progress in terms of "percent correct". There is no attention to sub-scores from which decisions might be made more precisely. Just as oral reading fluency measures are dependent on the quality or equivalence of the passages sets, curriculum sampling is dependent on the quality or equivalence of the alternate forms of the tests. Scores associated with assessing multiple skills can show significant variance, whereas short probes that assess one or two skills have more stable scores (Christ et al. 2008 summarized in Tindal 2013). Finally, there is no guarantee that the curriculum being sampled is the set of skills that each student needs. This final shortcoming is equally valid for MM if the MM is focused on a sequence of skills for one grade level—say the enrolled grade of the student. However, an MM that is focused on vertical strands of skills across grades is much more likely to hone in on a student's appropriate needs.

An additional issue for curriculum sampling in making decisions about the effectiveness of instruction is the issue of opportunity to learn. At the beginning of the year students will respond to probes on skills they have not been taught. It is hard to distinguish whether progress is based simply on exposure to skills or attributable to the quality of the instruction. In terms of student progress, this is moot; however, for judging the adequacy of an intervention, it is critical.

In MM, a skill is only measured for growth while the student is growing. The possibility of a student hitting a ceiling is central to the design of any MM system of measuring growth; skills are acknowledged as being constrained.

Curriculum-based measurement may not provide the measurement system that is ideal to best meet our objectives of being linked and sensitive to instructional change. (Shapiro 2013, 65)

Traditional View of Mastery Measurement: Low Rigor, Instructional Sensitivity

If we define instructional sensitivity as the ability of an assessment to measure changes in student knowledge and skills over short time durations, then MM, with their granularity and focus on what has just been instructed, are more instructionally sensitive than GOMs. Shapiro (2013, 65) notes this:

> *Curriculum-based measurement may not provide the measurement system that is ideal to best meet our objectives of being linked and sensitive to instructional change. Many schools routinely use metrics built into the instructional process, such as teacher-made tests, publisher-made tests…. However, perhaps there is a need to study and understand the potential of these types of less-than-rigorous measures in the decision-making processes for judging a student's response to intervention.*

Shapiro uses the phrase "less-than-rigorous" to refer to teacher- and publisher-made assessments that are built into the instructional process. MM—with its ever-changing probes "built into the instructional process"—provides new information with each administration. If these probes are given based on a well-defined framework of skills, information about what to teach next and when to move on is available whenever needed.

Given this, why has MM not been more widely adopted? MM, with its focus on making in-the-moment instructional decisions, had been deemed less useful for making high-stakes decisions. The Iris Center at Vanderbilt University states this clearly:

> *The remainder of this module will focus on CBM rather than MM, because CBM is an assessment procedure that is easily and quickly implemented and the results provide a clear, visual representation of how students are progressing academically.*

The notion of "easily and quickly" focuses on the idea that developing end-of-grade probes that reflect the skills for the year can be done quickly if the skills are limited in scope. Comparability of the probes, on the surface, can be achieved by matching similar items across probes. For MM, “quick and easy” has meant drawing a few items from various tests and quizzes, and has thus been seen as having low rigor and having no way to link probes across time. Without the ability to link probes to create a “visual representation” of how students are progressing in terms of the whole domain of interest, MM has not been deemed an adequate alternative for GOM. MMs, as typically implemented, are in-the-moment ad hoc assessments that do not necessarily depend on a completely defined framework of skills, but do provide information about the skill of interest.

Some of the reason comes from the perspective of what has been practical for districts or university researchers to develop or implement. For example, it is quite possible to pull a set of grade level reading passages for oral reading fluency; furthermore, it is possible to improve the quality and cohesion of those passages. Similarly, it is possible to develop a set of computation probes that cover all the computation skills in one grade and to develop alternative forms of that probe. What is beyond the scope of districts and even university researchers is to develop a fully articulated set of skills for each grade and then sequence those skills in a logical instructional order and, after that, develop rigorous probes for each of the skills. It is this high barrier to creating rigorous well-structured MMs, that has left the field seeing MMs as low rigor and having as many structures as there are curricula on which it is being implemented. As an example of the high barriers to creating a rigorous and structured MM system, NWEA™ took several years to develop Skills Navigator with more than 1000 skills and 10,000 items.

MMs, with their granularity and focus on what has just been instructed, are more instructionally sensitive than GOMs.

Research shows that GOMs are unable to support high-stakes decisions, and GOMs are less useful for predicting success beyond the early grades. A consideration of MMs that can provide both instructional utility and usefulness in the broader and more complex curricula in higher grades makes sense. What remains is to overcome the shortcomings of traditional MM applications.

A New Vision for Mastery Measurement: Rigor and Structure

To take full advantage of the formative information provided by MM, it is necessary to bring rigor and structure to how MM is used. A rigorous system would have probes that are consistently administered, that are well-aligned to skills, and that meet industry standards for item quality. Furthermore, the system would have clear rules for mastery and non-mastery, and apply those rules across all skills and grades. Such a rigorous system would consistently interpret and report probe results.

Structure comes from a predefined framework of skills that allow for clear formative decisions about what is next instructionally. Developing structure faces issues of both scope and sequence. Even in an era in which standards across many states are the same or variations of the same standards, the grade level curriculum will vary from classroom to classroom—both in terms of scope, exactly what is covered, and sequence, the order in which it covered. One way to develop a framework that addresses the scope issue is to limit the scope to skills that are essential to grade level mastery. That is focus on skills at each grade that are the building block skills on which a variety of curricula might be developed. A modular framework of groups of logically related skills would allow for a variety of instructional sequences within a grade, by matching the modules to the instructional sequence.

To address the issue of finding the right set of skills for each student, the framework needs a vertical as well as a grade level structure. The modular groups of skills from each grade would need to be sequenced into logical instructional strands. Such a structure would allow the development of probes designed to locate students within each strand of skills. Locating students within a vertical framework with one set of probes and then using rigorous probes to determine mastery would provide strong information to make formative decisions. Additionally, each skill should have sufficient mastery probes to address the issue of skill retention over time.

Locating students within a vertical framework with one set of probes and then using rigorous probes to determine mastery would provide strong information to make formative decisions.

Such an MM can provide educators with precise information to inform decisions about what to teach, what to teach next, and when to move on. The question remains whether an MM—based on the structure described and coupled with high-quality probes that both locate students and assess mastery—can also provide information to inform decisions about the effectiveness of instruction. That is, is a student "on track" to meet instructional goals. Rather than relying on slope, an MM based on a strand of skills would look at the proportion of skills mastered. An MM system would allow educators to define the skills that need to be mastered and to define a time frame for completion at a strand level. A bar graph that is used to represent the progress of instruction would display percent of skills mastered at any one time. Determining effectiveness of progress of skill mastery relies on the common-sense notions of completing a task: the proportion of skills mastered should be at least equal to the proportion of instructional time used (e. g. if 50% of the instructional time has passed, then at least 50% of the skills should be mastered to be "on track."). This view of progress is much more focused than the global look provided in GOMs—both in terms of scope of skills and timeframe. Whether this focused look provides information for making RTI decisions is a topic for further study.

Conclusion

Progress monitoring is characterized by repeated measurement, but research indicates that more frequent administration of GOMs does not provide better information for making important decisions. Rather, it is administration of a high-quality GOM over longer time periods that improves decision making. With research converging on durations of at least eight—and, more reliably, 12 to 14—weeks, the information for decision making can just as effectively come from interim assessments like MAP. Furthermore, "robust indicators," the most powerful of which is oral reading fluency, tend to be more useful for tracking progress in the lower grades, where the curriculum is narrower and less complex.

Given this, it makes sense to couple a high-quality interim assessment with more instructionally sensitive assessments that provide new actionable information with each administration. This is Shapiro's (2013) second recommendation. MAP and Skills Navigator provide this combination of a high-quality interim assessment that can be given every 12 to 15 weeks, with a rigorous and well-designed MM classroom assessment tool that provides formative information as frequently as desired.

References

Ardoin, Scott P., and Theodore J. Christ. 2009. "Curriculum-Based Measurement of Oral Reading: Standard Errors Associated with Progress Monitoring Outcomes from DIBELS, AIMSweb, and an Experimental Passage Set." *School Psychology Review* 38 (2): 266–83.

Ardoin, Scott P., Theodore J. Christ, Laura S. Morena, Damien C. Cormier, and David A. Klingbeil. 2013. "A Systematic Review and Summarization of the Recommendations and Research Surrounding Curriculum-Based Measurement of Oral Reading Fluency (CBM-R) Decision Rules." *Journal of School Psychology* 51 (1): 1–18. doi:10.1016/j.jsp.2012.09.004.

Christ, Theodore J. 2006. "Short-Term Estimates of Growth Using Curriculum-Based Measurement of Oral Reading Fluency: Estimating Standard Error of the Slope to Construct Confidence Intervals." *School Psychology Review* 35 (1): 128–33.

Christ, Theodore J., Barbara D. Monaghen, Cengiz Zopluoglu, and Ethan R. Van Norman. 2013. "Curriculum-Based Measurement of Oral Reading: Evaluation of Growth Estimates Derived with Pre-Post Assessment Methods." *Assessment for Effective Intervention* 38 (3): 139–53. doi:10.1177/1534508412456417.

Christ, Theodore J., Cengiz Zopluoglu, Jeffrey D. Long, and Barbara D. Monaghen. 2012. "Curriculum-Based Measurement of Oral Reading: Quality of Progress Monitoring Outcomes." *Exceptional Children* 78 (3): 356–73. doi:10.1177/001440291207800306.

Deno, Stanley L. 1985. "Curriculum-Based Measurement: The Emerging Alternative." *Exceptional Children* 52 (3): 219–32. doi:10.1177/001440298505200303.

Deno, Stanley L., and Phyllis K. Mirkin. 1977. "Data-Based Program Modification: A Manual." Guide. Washington, D.C.: Council for Exceptional Children. http://eric.ed.gov/?id=ED144270.

Foegen, Anne, Cynthia Jiban, and Stanley Deno. 2007. "Progress Monitoring Measures in Mathematics: A Review of the Literature." *Journal of Special Education* 41 (2): 121–39. http://files.eric.ed.gov/fulltext/EJ775117.pdf.

Fuchs, Douglas, Donald L. Compton, Lynn S. Fuchs, and Joan D. Bryant. 2006. "The Prevention and Identification of Reading Disability." Paper presented at the Pacific Coast Research Conference. San Diego: CA.

Fuchs, Lynn S. 2004. "The Past, Present, and Future of Curriculum-Based Measurement Research." *School Psychology Review* 33 (2): 188–92.

Hintze, John M., and Theodore J. Christ. 2004. "An Examination of Variability as a Function of Passage Variance in CBM Progress Monitoring." *School Psychology Review* 33 (2): 204–17.

Jenkins, Joseph R., J. Jason Graff, and Diana L. Miglioretti. 2009. "Estimating Reading Growth Using Intermittent CBM Progress Monitoring." *Exceptional Children* 75 (2): 151–63. doi:10.1177/001440290907500202.

Jenkins, Joseph R., J. Jason Graff, and Diana L. Miglioretti. 2006. "How Often Must We Measure to Estimate Oral Reading Growth?" Paper presented at the Pacific Coast Research Conference. San Diego: CA.

Jenkins, Joseph R., Roxanne F. Hudson, and Sung Hee Lee. 2007. “Using CBM-Reading Assessments to Monitor Reading Progress.” *Perspectives on Language and Literacy* 33 (2): 11–18.

Jenkins, Joseph, and Kari J. Terjeson. 2011. “Monitoring Reading Growth: Goal Setting, Measurement Frequency, and Aethods of Evaluation.” *Learning Disabilities Research & Practice* 26 (1): 28–35. doi:10.1111/j.1540-5826.2010.00322.x.

Ketterlin-Geller, Leanne R., Diane B. Gifford, and Lindsey Perry. 2015. “Measuring Middle School Students’ Algebra Readiness: Examining Validity Evidence for Three Experimental Measures.” *Assessment for Effective Intervention* 41 (1): 28–40. doi:10.1177/1534508415586545.

Paris, Scott G. 2005. “Reinterpreting the Development of Reading Skills.” *Reading Research Quarterly* 40 (2): 184–202. doi:10.1598/RRQ.40.2.3.

Shapiro, Edward S. 2013. “Commentary on Progress Monitoring with CBM-R and Decision Making: Problems Found and Looking for Solutions.” *Journal of School Psychology* 51 (1): 59–66. doi:10.1016/j.jsp.2012.11.003.

Shapiro, Edward S., Alexandra Hilt-Panahon, Karen L. Gischlar, Kathleen Semeniak, Erin Leichman, and Shelly Bowles. 2012. “An Analysis of Consistency between Team Decisions and Reading Assessment Data within an RTI Model.” *Remedial and Special Education* 33 (6): 335–47. doi:10.1177/0741932510397763.

Silberglitt, Benjamin, David Parker, and Paul Muyskens. 2016. “Assessment: Periodic Assessment to Monitor Progress.” In *Handbook of Response to Intervention*, 271–91. Boston, MA: Springer US. doi:10.1007/978-1-4899-7568-3_16.

Sisco-Taylor, Dennis, Wenson Fung, and H. Lee Swanson. 2015. “Do Curriculum-Based Measures Predict Performance on Word-Problem-Solving Measures?” *Assessment for Effective Intervention* 40 (3): 131–42. doi:10.1177/1534508414556504.

Tindal, Gerald. 2013. “Curriculum-Based Measurement: A Brief History of Nearly Everything from the 1970s to the Present.” *ISRN Education* 2013: 1–29. doi:10.1155/2013/958530.

Wayman, Miya Miura, Teri Wallace, Hilda Ives Wiley, Renáta Tichá, and Christine A. Espin. 2007. “Literature Synthesis on Curriculum-Based Measurement in Reading.” *Journal of Special Education* 41 (2): 85–120. doi:10.1177/00224669070410020401.