Progress Monitoring for Accelerating the Pace of Learning:

Considerations and Recommendations for Selecting Academic Measures

FastBridge Learning

Table of Contents

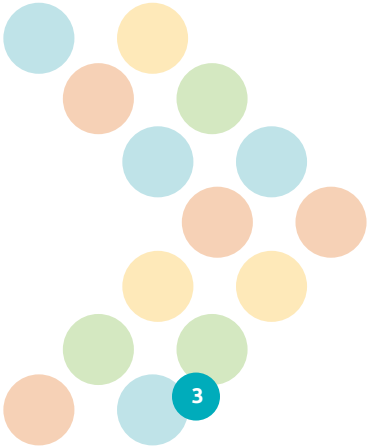
Importance of Progress Monitoring

Data-based decision making is a key component of Response to Intervention (RTI) and Multi-tiered System of Support (MTSS) initiatives. Such systems employ universal screening in order to identify all students' learning needs. Those students whose skills are below expectations (i.e., benchmarks) are provided with evidence-based interventions. Educators then use progress monitoring data to evaluate whether a student's rate of improvement (ROI) is indicating an adequate response to the intervention. This information is then used to make decisions about whether instructional practices and interventions should be maintained, modified, or intensified in order to ensure that students are receiving supports and instruction to propel their learning and match their needs. Reliable and valid progress monitoring measures are an even more important topic now that more states and districts are utilizing RTI data as part of the evaluation procedures for higher stakes decisions such as identifying students with specific learning disabilities (IDEIA, 2004). The goal of this paper is to describe the purpose of progress monitoring, to present considerations for selecting effective and appropriate progress monitoring measures, and to provide a comparison of two assessment approaches for progress monitoring.

"Using progress monitoring to guide instruction has benefits: students learn more, teacher decision making improves, and students are more aware of their performance."

Purpose and Use of Progress Monitoring

Progress monitoring is defined as the frequent assessment of students who are participating in supplemental intervention. It can inform whether the addition of the intervention is helping the student to achieve the intervention goal. Monitoring student progress involves administering measures repeatedly to estimate students' growth in a specified skill. Progress monitoring is used to assess students' academic progress, to examine rate of improvement, and to evaluate effectiveness of instruction or intervention. Progress monitoring is typically used with both individual students and small groups. In addition to monitoring student growth, research evidence suggests that using progress monitoring to guide instruction has benefits: students learn more, teacher decision making improves, and students are more aware of their performance (Kovaleski, 2007; NCRTI, 2018; Shinn, 2008).



Core Features of Progress Monitoring Measures

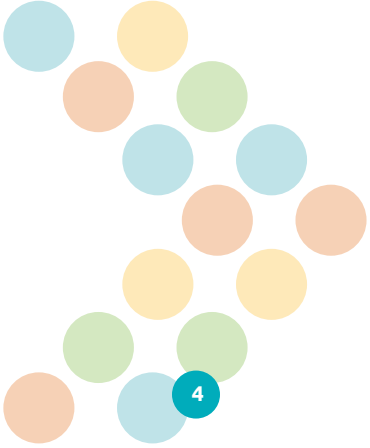
Educators and school administrators are faced with the important decision of selecting which progress monitoring measure to use for their students. There are a number of key considerations for this decision.

First, progress monitoring measures should be **brief and simple.** Assessment should not consume too much valuable instruction time. This is especially important because progress monitoring measures are typically administered at frequent, regular intervals (e.g., weekly or every other week). The longer that it takes to administer and score an assessment, the less time there is available for instruction. In addition to being brief, measures must be simple. If teachers do not understand how to administer or interpret the data, they will likely not be able to use the data to modify their instruction.

"Measures should be brief and simple, sensitive to growth, and reliable and valid."

Second, measures must be **sensitive to growth** to be able to show the effects of the intervention over short periods of time. Normative growth rates (i.e., research-based recommendations on growth) and decision rules for interpreting progress data should be available so that teachers can determine whether or not a student is responding to intervention. For example, if a measure has a growth rate of 1 item per month, that would be .25 items or points per week which is difficult to interpret in relation to instructional decisions. In other words, how does one know if a student answers a fourth of a question correctly? When a measure has more room for growth (e.g., 8 items or points per month) growth – or lack thereof – is more easily understood and instructional decisions can be made more promptly and reliably.

Third, measures should have technical evidence of **validity and reliability.** Validity refers to the extent that an assessment has evidence that it measures what it purports to measure for the intended use (e.g., a math measure is measuring math progress rather than reading ability). Students typically receive interventions that target a specific area of need. It is important that the progress monitoring measure matches the student's learning goal. For instance, if a student has difficultly with fluently adding math facts, the best intervention would likely focus on fluency and be monitored with an addition fluency progress measure. If that student were monitored with a multiplication, or geometry progress measure, the data would not yield accurate results about the student's response to that particular intervention.



Reliability refers to the consistency of the measure over time or across different forms and determines the extent to which we can depend on the accuracy of the data (i.e., does the score actually reflect what the student can do). Without reliability and validity evidence for a progress measure, it is unclear whether the data provide information that actually tells educators the truth about student performance. Since progress monitoring is conducted over time, it is important that the assessment is valid for the target skills over time and reliable and predictive of later outcomes.

Types of Progress Monitoring Measures

The most widely researched and commonly used progress monitoring assessments are curriculum-based measures (CBMs). CBMs were first developed for the purpose of measuring student growth. CBMs have also historically been used for making decisions about screening, referrals, program outcomes and IEP outcomes. Another type of academic measure commonly use in schools today is computer adaptive testing (CAT). CATs were originally developed for the purpose of replacing traditional fixed-length paper-and-pencil tests of achievement. Recently, some schools have begun to use CATs to monitor individual student progress; however, the evidence for this use is weak. FastBridge Learning® has a suite of CBMs and CATs to measure math and reading skills. While CATs have demonstrated strong evidence for screening, there is very limited research concerning their use for progress monitoring. FastBridge Learning recommends CBMs for monitoring an individual student's response to intervention because this use is well-supported by the available research. It is important for educators to understand the distinctions between CBM and CAT in order to make the best selections for progress monitoring measures. In the following section, CBMs and CATs are described and the two assessment approaches are compared in the context of monitoring student progress.

Curriculum-Based Measures

CBMs were originally developed using samples from classroom teaching materials in order to provide assessments directly reflecting curricula. The development of CBM was intended to provide a brief, repeatable, authentic and inexpensive measure to track student progress (Deno, 1985; Fuchs, Fuchs & Hamlett, 1990). It is worth noting that CBM typically incorporates standardized procedures for administration and scoring. Various distributors of CBM develop specific procedures for administration and scoring that are specific to their set of CBM materials. Each publisher also provides guidelines for interpretation and use, which often include a specific set of standardized benchmarks and norms.

CBMs are often timed because standardized assessments completed under timed conditions provide evidence of a student's automaticity, or fluency, with the target skill. In addition, measuring fluency enhances a measure's sensitivity to instruction so that it can identify small increments of student achievement. The most robust CBM research addresses CBM of oral reading fluency (CBM-R). There are also CBMs that assess the early reading skills of phonemic awareness and phonics. More recently CBM of early math and mathematics computation have been developed.

Early research about progress monitoring identified CBM-R as an ideal measure given its strong relation to broad reading, its brevity in administration (1 minute) and the ability to match assessment to the targeted outcomes of specific interventions. The research about CBM-R for progress monitoring is extensive and spans over 30 years. Much of the early research was in relation to the measure's reliability and validity as well as its ability to identify students at-risk for difficulties (Deno, 1985; Reschly et al., 2009, Wayman et al., 2007). Additional research examined student growth rates and sensitivity to the effects of instruction (Deno et al., 2001; Hintze, Shapiro & Lutz, 1994; Shinn, Gleason & Tindal, 1989). Notably, results show that when CBM is used it results in increased rates of student improvement (Fuchs, Fuchs, & Stecker, 1989; Stecker, Fuchs & Fuchs., 2005). This finding has been associated with the "feedback" loop provided to the student and teacher when progress data are reviewed on a regular basis. In addition, CBM was found to be sensitive to group-level growth as a function of the type of instruction over brief assessment periods. Based on the accumulated evidence, CBM is considered ideal for progress monitoring.

"Based on the accumulated evidence, CBM is considered ideal for progress monitoring."

Computer-Adaptive Tests

CATs were developed as an alternative to fixed item achievement tests and have been proven to be a helpful measure to identify each student's achievement levels in reading and mathematics (Weiss & Kingsbury, 1984). CATs utilize item response theory (IRT) and use student answers in real time to inform subsequent questions based on difficulty level. IRT is a statistical method that calculates the difficulty of all questions in a "bank" of testing items and then uses selected items in relation to each student's response pattern. Specifically, when taking a CAT, the student starts with items matched to grade level, but later items are selected by the computer program based on the student's answers to earlier items (Weiss, 2004; Weiss & Kingsbury, 1984). For example, a third-grade student starts with several third grade questions but then the items get easier or harder based on whether the first items were answered correctly. In this sense, CATs automatically adjust to a student's skill level to measure broad achievement. CATs typically take between 15 and 60 minutes to administer (NCRTI, 2018) and

are primarily used in educational settings for the purposes of screening and identifying students at risk for difficulties.

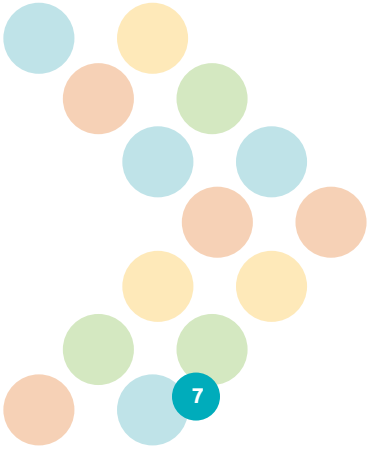
Although originally designed for measuring broad achievement (i.e., screening), many educators have wondered whether CATs can be used for progress monitoring as well. While some research has begun to emerge to provide potential evidence for CATs to measure group progress (Shapiro, Dennis & Fu, 2015; Shapiro & Gibbs, 2014), the research base is not currently strong enough to support the use of typical CATs for frequent progress monitoring. A review of the research is provided below in the context of the key features of effective progress monitoring measures.

Comparison of the Research for Progress Monitoring

As with all educational decisions there are advantages and disadvantages to selecting different assessment approaches. When considering assessments for a specific purpose (e.g., progress monitoring), educators need to evaluate whether the available options have research supporting their use for that purpose. This means that research evidence that a measure is reliable and valid for screening does not mean it has the same properties for progress monitoring (and vice versa). It is vital for decision makers to use caution that they do not use only evidence of screening to select measures for regular progress monitoring. The following section presents a comparison of research evidence of both CBM and CAT in regard to progress monitoring. Some educators may prioritize different features based on their district goals and needs. These advantages and disadvantages are discussed in relation to the core features of progress monitoring measures discussed earlier: Measures should be brief and simple, sensitive to growth, and reliable and valid.

Brief and Simple

CBMs have the advantage of time. CBMs are typically timed and require less time to administer than CATs. For example, CBMReading takes 1 minute to administer. CATs are generally untimed and take at least 15 minutes or often longer to complete. One of the reasons that CBMs are brief is that they focus on a small number of very specific skills. The simplicity of CBM offers the advantage of parallel forms so that items are uniform across students. These forms consist of the types of reading passages or math problems using different words or letters, problems or numbers so that the version that a student completes one week is different yet measuring similar skills the following week. By comparison, in order to adapt and provide questions matched to each student's skill level, CATs include items reflecting a broad range of skills. Although the items included on each CAT session are specific to the student, because the items vary by student it is difficult for educators to understand instructional information.



Sensitive to Growth

Sensitivity to growth is vital in progress monitoring measures because the purpose is to accurately detect improvement, or lack thereof, during brief periods of time. This feature of progress monitoring measures is critical in order for growth data to document whether an intervention meets a student's learning needs. When progress data are used for making a decision about disability identification (e.g., specific learning disability) measure sensitivity is imperative. A vast amount of CBM research demonstrates its utility to estimate linear growth over time for student groups, including evidence that CBMs are sensitive to changes in instruction and vary across groups of students (e.g., students eligible for special education; Deno, 1985; Deno et al., 2001; Hintze et al., 1994; Shinn, Gleason & Tindal, 1986). Another consideration in regard to growth is the evidence for decision rules. In other words, how will educators know when a student has or has not made growth. The interpretation of CBM data utilizes a research method known as single-case design. There is a strong research base that supports the accuracy of SCD decision (Johnston & Pennypacker, 2008; Kazdin, 2010; Newell & Christ, 2017; Van Norman, Christ & Newell, 2017). This is important in refining decisions made by human judgment and reducing test error. In contrast to CBM research, there is very limited research about decision rules for interpreting individual student growth using CATs.

Compared to the research on CBMs, the research on sensitivity of growth for CATs is in its infancy. Some of the original research included a technique to measure individual change (Weiss & Kingsbury, 1984), however this method is not yet used in current applications of CAT in educational settings. A recent study compared a common CAT and CBM in the area of reading for two groups of elementary students (Shapiro & Gibbs, 2014). Findings indicated that CATs do show growth over the course of a year. That said, CAT progress monitoring research has yet to demonstrate specific benefits for selecting student-specific interventions and measuring individual growth.

Another recent finding in regard to CATs and progress monitoring is that individual growth estimates vary based on the frequency of data collection (Nelson, Van Norman, Klingbeil & Parker, 2017). Specifically, observed growth when using 5 data points was significantly different than when using 2 data points across 18 weeks. Nelson et al. found that variability and probable error were high enough that the number of data points led to different educational decisions. In regard to CBMs, previous progress monitoring research (e.g., Christ, Zopluoglu, Monaghen & Van Norman, 2013; Thornblad & Christ, 2014) also showed that the number of data points has a significant role in how data are interpreted. Yet, recent findings about improved calculation methods for CBM progress data (Christ & Desjardins, 2018) trend lines suggest that those prior limitations have been overcome for CBM-based progress data. At this time, available research confirms that CBMs are sensitive to student growth when sufficient data are collected. There are too few studies of CAT for progress monitoring to draw firm conclusions about sensitivity to growth over time.

Reliability and Validity

A measure is only useful if it demonstrates reliability and validity for the intended use. Both CBMs and CATs have demonstrated reliability and validity for a specific time point when used for screening, and there is substantial evidence of CBMs' reliability and validity for progress monitoring. In regard to CBMs, experts suggest that CBM-R outcomes are reliable, valid and precise when data are collected often and for an extended period of time (Christ, Zopluoglu, Long & Monaghen, 2012; Christ et al., 2013; Thornblad & Christ, 2014) or when improved calculation methods are used (Christ & Desjardins, 2018). Ideal reliability and validity is reached when a very good progress monitoring measure is used with a minimum of 12- 14 data points over at least 6 weeks. While this may seem like a lot of data, it is important to remember that the brevity of each CBM administered will require only 15 minutes to collect those data points. This is in strong contrast to administration time for multiple administrations of CATs which would take hours over a similar time period.

"Additional research is needed to confirm whether CAT progress data are psychometrically sound for evaluating student progress."

Additional research is needed to confirm whether CAT progress data are psychometrically sound for evaluating student progress. Although Shapiro et al. (2015) compared growth estimates from monthly administrations of a CAT in math with two common math CBM measures among elementary aged students, their findings did not establish whether the CAT data provided truly unique information related to student performance. When controlling for CBM growth, the CAT growth explained additional information on a state achievement test over the course of a school year. While this provides some evidence of possible predictive validity in addition to CBM, it does not provide evidence for the predictive validity of CAT as a progress measure on its own. Additionally, the decisions made from progress data usually focus on a student's need for additional services (e.g., interventions or special education) and not in relation to performance on state tests.

Content

Another distinction between CBM and CAT is in relation to the scope of content assessed. CBMs focus on a small number of specific skills, but CATs usually include a broad range of items across multiple scales (e.g., vocabulary, comprehension). Some educators argue that CBM is too narrow and that CAT is beneficial because it measures all domains. Usually, not all domains are important for progress monitoring, because interventions focus on specific skills. Although some CATs provide subscale "domain" scores in their reports,

these specific domain scores provided by some CAT assessments do not have strong evidence of reliability and validity (FastBridge Learning, 2018). A strong benefit of CBM as a progress measure is that it measures specific skills while requiring less time to administer and score. While the use of a broad measure may sound appealing, educators need to consider the intended purposes of different assessments and compare the benefits of broader assessment with the amount of time required for tests that include more skill areas.

Conclusions

Expert educators and researchers who collaborate with FastBridge Learning currently recommend the use of CBMs as the ideal progress measure because they have the strongest research evidence at this time. We believe that educators should focus their time on instruction and progress monitoring assessments should be as brief as possible while still yielding reliable and valid results. We also believe that ongoing professional development is important so that educators can be aware of the advantages and limitations of different progress measures. In general, decisions about student response to intervention should never be made using one source of data and multiple data sources (e.g., implementation fidelity, student engagement in intervention, student attendance, other assessments, etc.) should be considered on a case-by-case basis to make the best educational decisions. When used appropriately, CBMs have the best evidence base for determining student progress.

This way to FASTER results.

For more information about formative assessment, visit **www.fastbridge.org**.

References

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C. & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding Curriculum-Based Measurement of oral reading fluency (CBM-R) decision rulues, *Journal of School Psychology,* 51, 1 – 18.

Christ, T.J. & Desjardins, C. D. (2018). Curriculum-based measurement of reading: an evaluation of frequentist and Bayesian methods to model progress monitoring data. *Journal of Psychoeducational Assessment,* 36, 55–73

Christ, T. J., Zopluoglu, C., Long, J. D., & Monaghen, B. D. (2012). Curriculum-based measurement of oral reading: Quality of progress monitoring outcomes. *Exceptional Children,* 78, 356-373.

Christ, T. J., Zopluoglu, C., Monaghen, B. D., & Van Norman, E. R. (2013). Curriculum- Based Measurement of Oral Reading: Multi-study evaluation of schedule, duration, and dataset quality on progress monitoring outcomes. *Journal of School Psychology,* 51, 19–57.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children,* 52, 219-232.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review,* 30, 507–524.

Fuchs, L. S., Fuchs, D., & Stecker, P. M. (1989). Effects of curriculum-based measurement on teachers' instructional planning. *Journal of Learning Disabilities,* 22, 51–59.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1990). Curriculum-based measurement: A standardized long-term goal approach to monitoring student progress. *Academic Therapy,* 25, 615–632.

Hintze, J. M., Shapiro, E. S., & Lutz, J. G. (1994). The effects of curriculum on the sensitivity of curriculum-based measurement in reading. *The Journal of Special Education,* 28, 188–202.

Hintze, J. M., Wells, C. S., Marcotte, A. M., & Solomon, B. G. (2018). Decision-making accuracy of CBM progress-monitoring data. *Journal of Psychoeducational Assessment,* 36, 74-81.

Individuals with Disabilities Education Improvement Act of 2004 (2004). PL 108–446, 20 U.s.C §§ 1400 et seq. : U.S. Department of Education.

Johnston, J. M., & Pennypacker, H. S. (2008). *Strategies and tactics of behavioral research,* 3rd ed. New York: Routledge.

Kazdin, A. (2010). *Single-case research designs: Methods for clinical and applied settings.* London: Oxford University Press.

Mellard, D., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research & Practice,* 24, 186–195.

National Center for Response to Intervention (NCRTI). (2018). *Screening Tools Chart Rating System.* Retrieved from: http://www.rti4success.org/resources/tools-charts/screening-tools-chart

Nelson, P. M., Van Norman, E. R., Klingbeil, D. A., & Parker, D. C. (2017). Progress monitoring with computer adaptive assessments: The impact of data collection schedule on growth estimates. *Psychology in the Schools,* 54(5), 463-471.

Newell, K. W. & Christ, T. J. (2017). Novice interpretations of progress monitoring graphs: Extreme values and graphical aids. *Assessment for Effective Intervention,* 42, 224 – 236.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology,* 47, 427-469.

Shapiro, E. S., Dennis, M. S., & Fu, Q. (2015). Comparing computer adaptive and curriculum-based measures of math in progress monitoring. *School Psychology Quarterly,* 30, 470-487.

Shapiro, E. S., & Gibbs, D. P. (2014). *Comparison of progress monitoring with computer adaptive tests and curriculum-based measures.* Bethlehem, PA: Center for Promoting Research to Practice, Lehigh University. Available online from http://coe.lehigh.edu/cprp/research/current

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., Gleason, M. M., & Tindal, G. (1989). Varying the difficulty of testing materials: Implications for curriculum-based measurement. *Journal of Special Education,* 23, 223–233.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools,* 42, 795–819.

Thornblad, S. C., & Christ, T. J. (2014). Curriculum-based measurement of reading: Is 6 weeks of daily progress monitoring enough? *School Psychology Review,* 43, 19-29.

Van Norman, E. R., Christ, T. J., & Newell, K. W. (2017) Curriculum-based measurement of reading progress monitoring: The importance of growth magnitude and goal setting in decision making. *School Psychology Review,* 46, 320-328.

Van Norman, E. R., Nelson, P. M., & Parker, D. C. (2017). Technical adequacy of growth estimates from a computer adaptive test: Implications for progress monitoring. *School Psychology Quarterly,* 32(3), 379.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education,* 41, 85–120.

Weiss, D. J. (2004). Computerized Adaptive Testing for Effective and Efficient Measurement in Counseling and Education, *Measurement and Evaluation in Counseling and Development,* 37, 70-84.

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement,* 21, 361-375.

FastBridge Learning Resources

Arañas, Y. (2015, December 16). Progress Monitoring Basics. [Blog post]. Retrieved from http://www.fastbridge.org/2015/12/progress-monitoring-basics/

Arañas, Y. (2015, December 21). Using Progress Monitoring Data to Make Instructional Decisions. [Blog Post]. Retrieved from http://www.fastbridge.org/2015/12/using-progress-monitoring-data-to-make-instructional-decisions/

FastBridge Learning. (2016). FASTfact: Understanding strand score limitations and suggestions. Retrieved from: https://fastforteachers.freshdesk.com/support/solutions/articles/5000725628-fastf-act-understanding-strand-scores-limitations-and-suggestions

McKevett, N. M. (2017, October 13) Conducting Progress Monitoring. [Blog post]. Retrieved from: http://www.fastbridge.org/2017/10/conducting-progress-monitoring/