Progress Monitoring in Middle School Mathematics

Crysti Jones

Vanderbilt University

Abstract

As education moves towards a more data-based approach, it's important that all educators have access to reliable and valid methods of collecting data on their students' progress. The purpose of this paper is to explore available progress monitoring measures at the secondary level. According to Foegen, Jiban, and Deno (2007), progress monitoring research focuses on three stages: technical adequacy, student growth, and instructional utility. I focused my search on papers that targeted these three stages.

In order to identify appropriate materials, I searched 44 databases in PsychInfo for articles in peer-reviewed, scholarly using the search terms: meta-analysis, literature review, algebra, progress monitoring, curriculum-based measurement, math, and secondary education. From here I attempted to identify three groups of articles. The first group was comprised of meta-analyses or literature reviews of research done on progress monitoring. I found six papers that met these requirements, but only included the two that discussed secondary measures.

The next group I created focused on secondary measures. Since this area is still relatively new, these papers focused on technical adequacy and student growth. I focused on measures at the secondary level and identified three papers by the same researchers regarding the development of a set of algebra measures for general and special education students.

Finally, I identified a group that was focused on implementation and use of procedures in the classroom. I found one article related to the group of algebra measures and one article related to general use of math progress monitoring tools to improve student development.

Introduction

What is progress monitoring?

Progress monitoring is the act of monitoring how well students are mastering academic skills. One form of progress monitoring is curriculum-based measurement. Curriculum-based measurements are a set of standardized, research-validated procedures used to identify and analyze the level and trend of student's performance and/or growth in reading, math, writing, or spelling. They are used to screen students for academic problems, identify appropriate placement, evaluate student's response to instruction, evaluate program effectiveness, and inform teachers on instructional decisions (Christ, Scullin, Tolbize, & Jiban, 2008; Stecker, Fuchs, & Fuchs, 2005). These measurements are frequent, brief (1- to 5-min), empirically documented for reliability and criterion validity, and provide teachers with objective data on student performance (Foegen, 2008).

Teachers can use these forms to help determine a child's present level of performance and identify a target goal in order to set a trend line. As the year progresses, teachers conduct regular assessment. The frequency of assessment is determined by student needs and may range from weekly, biweekly, monthly, or more. Student data is then graphed and compared to the trend line (Stecker, Fuchs, & Fuchs, 2005). Students who continue to fall below their trend line may require additional support or instructional changes. Students who fall above the trend line should be given a new goal based on previous progress.

Review of Studies

Group One: Literature Review

Foegen, Jiban, and Deno (2007).

The purpose of this article was to review existing empirical literature on math progress monitoring measures. They identified two approaches for developing CBM in Math (CBM-M): curriculum sampling and robust indicators. Curriculum sampling is directly linked to the curriculum and can be used for diagnostic feedback; however, it can only measure one-year's worth of growth. Robust indicators measure broad math proficiency and can be used across grade levels, but are less useful to teachers for diagnostic purposes. They also identified three stages of research that they used to categorize articles: explores the technical adequacy of CBM use as static indicators, continuous progress monitoring, slope, variability, and student growth, and instructional utility of CBM.

They found that a majority of CBM-M studies focused on elementary and stage one research. Reliability of most forms were .8 and covered internal consistency, test-retest, alternate form, generalizability, and dependability. Validity measures were generally .8 to .9. Only a handful of studies looked at student growth, which varies greatly between measures. There were only four articles that looked at secondary measures, all conducted since 2000 and didn't include high school (as of 2007). Stage 1 research identified secondary CBM-Ms for basic facts, estimation, and conceptual understanding. Reliability for middle school was .8 or higher, which matched averages for all grades. Validity, criterion, was .29 to .66, most of .4 to .5 which was lower than averages for more well-researched elementary measures. Validity with state math tests was .8, which matches average validity for CBM. Stage 2 research looking at student growth focused on concept measures and their increase over four administrations. Most research didn't identify significant growth for secondary measures. Stage 3 identified seven studies that used static goals and dynamic goals, dynamic goals resulted in higher student growth. They also found that technology enhanced student performance.

The authors of this study concluded that secondary research isn't sufficient to make strong recommendations for use and dependability of CBM-M. They suggest that teachers consider the pros and cons of curriculum sampling and robust indicators, degree to which curriculum is represented by the measure, and the rate of student growth on specific measures before selected CBM-M to use in their classroom.

Christ, Scullin, Tolbize, and Jiban (2008).

This article focused on recent research in CBM-M until 2008. They identified four types of CBM-M: subskill mastery measures, general outcome measures, curriculum sampling, and robust indicators. Subskill mastery measures assess a single-skill over a brief period. General outcome measures assess a broad range of skills that are developed over the year. They also identified similar ranges of validity and reliability as Foegen, Jiban, and Deno (2007).

Their study found that context, duration, scope, format, and relative motivation all influence results. Difficulty often varies by multiple grades between forms and that teachers should use multiple measures to identify a median score. Single-skill CBM-M, subskill measures, can be used for low-stakes norm-referenced screening decisions, but aren't as dependable as multiple-skills. Multiple-skills measures provide greater variability and as such can be used to make high-stakes, criterion-referenced decisions.

They concluded that teachers should select measures based on context, duration, scope, format, and student motivation. Measures should systematically sample curriculum. The sequence of items should be random so that if students don't finish the entire probe, they will still have completed problems from each skill area. Teachers should administer probes for 4 to 14-minutes across multiple skills. When interpreting CBM-M teachers should consider the method.

Group Two: Middle School Math Measures

Foegen (2008).

This was a predominantly stage one and two study of middle school (MS) CBM-M for Project AAIMS. Foegen attempted to answer three questions: what is the reliability of MS measures, what is the criterion validity of measures, and how much academic growth do students display. They studied the use of Monitoring of Basic Skills Progress (MBSP) for basic math computation and concepts and applications, basic facts, estimation, complex quantity discrimination, and missing numbers. Two districts of seven general education teachers for grades 6^{th}, 7^{th}, and 8^{th} grade were randomly assigned to four conditions: a) MBSP Computation and estimation, b) MBSP concept application and basic facts, c) Estimation and complex quantity and missing numbers. All groups monitored students 4 times a year.

Foegen concluded that MBSP basic math computation had the largest standard deviations (SD), more variability in student scores, and greatest student growth rate. It had strong criterion validity and acceptable reliability. MBSP concepts and applications had the second largest SD, greater student change in performance, strong alternate form reliability, highest criterion validity at grades 6 and 7^{th}, greatest effect size, and highest predictive validity. Basic facts had high alternative form reliability, but lower growth rate than found in previous studies. Estimation had acceptable reliability on alternative forms, but lowest test-retest reliability. Complex quantity discrimination (a new measure) saw greater change in student performance, high alternative form reliability, highest criterion validity for 8^{th} grade, and the 2^{nd} largest effect size. Missing numbers (another new measure) had the smallest means and SD and acceptable alternative form reliability, overall it was the poorest performing measure.

The results of this study indicate that MBSP and Complex quantity discrimination scores were more consistent across time and forms and more related to external criteria and may make them good measure for teachers to use. Since this was one of the first studies at this grade, it would be difficult to make strong recommendations for use. The districts weren't very diverse and possibly not generalizable to other populations. While the research identifies a few potential MS measures, additional research would be required before any conclusions could be drawn.

Foegen, Olson, & Impecoven-Lind (2008).

This was a follow-up study attempting to refine old measures and test new math measures. This study identified robust indicators (basic skills, algebra concepts) and curriculum sampling (content analysis). This paper covered the progression that the researchers went through to develop the measures.

Initial use of measures was in in (2004) where researchers used 8, 8^{th} grade algebra classrooms ranging from advanced to special education. Classes were all given 2 forms of the basic skills (BS) and algebra concept (AC) measures, and one form of the content analysis (CA). Results indicate that BS and AC measures were scored with a high level of consistency, produced higher reliability coefficients and had strong relations with district math tests, student scores were also more significant across groups. CA was more difficult to score and had lower levels of reliability, making it potentially ineffective.

Additional revisions were made to determine how test-timing affected results. BS was changed to 2-, 3-, and 4-min probes. AC was changed to 4-, 5-, and 6-min probes. They also included two new measures for translations of conceptual understanding and content analysis – multiple choice (CA-MC). They conducted two studies with these measures. The first study looked at one district using the two new measures and the second district using all five (BS, AC, CA, Translation, and CA-MC). They found lower reliability and criterion validity than previous studies. Translation didn't seem to be an adequate measure of student knowledge, but the AC-MC was more reliable than previous measures.

Foegen (2008).

This study focused on the use of BS, AC, Translations, and CA-MC in both general and special education setting. This data was gathered from the previous mentioned study. Concurrent validity and predictive validity were also measured. Results indicated that concurrent validity was moderate (.4 to .6), but these results were stronger for students with disabilities. Predictive validity varied between schools and was relatively non-conclusive. Student growth for students with IEPs was lower on AC, but acceptable for BS and CA ranging from .5 to .79 points a week for BS, .3 for CA, and .24 to .37 for AC. While additional research is needed, initial findings indicate that these measures are sufficiently technically adequate for use with students with special needs. It seems that measures of BS and CA would be more appropriate than AC.

Group Three: Teacher Use and Student Growth

Foegen, Olson, and Impecoven-Lind (2008)

Part of their development of algebra measures looked at measuring teachers' use of data to make instructional changes. They used the same groups as their previous studies. Four parallel forms were made for BS, AC, and CA-MC. Teachers in two districts were broken into two groups. One was provided monthly feedback on student growth and suggestions for changes, the other was not. They created a spreadsheet template with individual graphs and a class-wide report. Research assistants scored student data and met with teachers on a monthly basis to go over results and make suggestions. Results indicated little difference between groups. Researchers believe this was because teachers were given data almost a month after it was relevant and as such teachers couldn't use it to make instructional changes. They conducted a second study where teachers were given control of scoring data and plotting student growth. Analysis of this data wasn't present at the time of publication, but initial findings indicate that teachers much preferred access and scoring their own data and potentially were able to use it successfully to make changes. These studies indicate that teachers having access to data isn't sufficient for producing changes in student growth.

Stecker, Fuchs, and Fuchs (2005).

This paper reviewed almost twenty years of research on helping change student growth through teaching practices. They reviewed dozens of studies from 1984 to 2008 in both general and special education classrooms. These studies looked at teacher use of data, research assistant supported use of data, and technology-support use of data to make instructional changes. They found that teachers, even in special education, make infrequent changes to instruction and that only when instructional changes were made to match student need, did student growth increase. Even when teachers were given the data, they often didn't understand how to use it to make appropriate changes. Stecker, Fuchs, and Fuchs state that frequent progress monitoring alone isn't sufficient for student growth. This paper makes five recommendations: a) CBM data must guide instruction, b) teachers must adhere to a data-decision framework where they both increase goals and adjust instruction based on student progress, c) computer applications can enable teachers to use data more effectively, d) skills analysis and coaching may help teachers identify students' areas of needs, and e) ongoing consultations are needed in order for teachers to make meaningful instructional changes.

Conclusion

Implications for the Classroom

Based on the current research, CBM-M are technically adequate for monitoring student growth on both curriculum sampled items and robust indicators. Teachers should know the difference between robust indicators and curriculum-sampling and which ones can be used to make instructional decisions (curriculum-based) and diagnostic decisions (robust indicators). Secondary math teachers should be aware of what research says is effective for use in middle schools. Project AAIMS has identified measures that are technically adequate and can effectively measure student growth on a weekly basis. While Project AAIMS makes all of the researched probes available, based on currently available research the most technically adequate measures for secondary math are Basic Math, Math Concepts and Algebra Concepts.

Teachers and administrators should also note that progress monitoring on a regular basis isn't sufficient for increasing student growth. Teachers need to be able to make instructional changes based off the data that the measures produce. Project AAIMS includes colored overlays that may help teachers identify strengths and weaknesses; however, none of their published research indicates whether or not teachers were able to use these to make instructional changes that resulted in added student growth. Administrators should be aware that research suggests the use of technology or coaching to help guide teachers in making effective changes.

Implications for Future Research

While there are probes available through Algebra, very little has been done to identify measures beyond this level. Researchers should continue to pursue stage 1 and stage 2 research to identify measures that are technically adequate and that are capable of measuring adequate levels of student growth. They should also pursue additional research on the use of these CBM-M to increase student growth. While Foegen, Olson, and Impecoven-Lind (2008) attempted to identify means of increasing student growth through data-analysis and instructional changes, they were unable to do so. One of the complaints made by teachers was that data wasn't produced fast enough to make instructional changes. Follow-up studies gave teachers immediate access to data, but it was unclear whether or not this helped. I wonder if using CBM-M data at this level will be significantly harder because teachers tend to spend significantly less time on concepts. Since teachers have so much material to cover, weekly progress monitoring may not be sufficient if teachers don't have time to re-teach old material. This may be an area that needs to be looked into further. The End ☺ .

REFFERENCES

Christ, T. J., Scullin, S., Tolbize, A., & Jiban, C. L. (2008). Implications of recent research: Curriculum-based measurement of math computation. *Assessment for Effective Intervention, 33*. doi:http://dx.doi.org/10.1177/1534508407313480

Foegen, A. (2008). Algebra progress monitoring and interventions for students with learning disabilities. *Learning Disability Quarterly, 31*. Retrieved from http://search.proquest.com.proxy.library.vanderbilt.edu

Foegen, A. (2008). Progress monitoring in middle school mathematics: Options and issues. *Remedial and Special Education, 29*. Retrieved from http://search.proquest.com.proxy.library.vanderbilt.edu

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education, 41*. Retrieved from http://search.proquest.com.proxy.library.vanderbilt.edu/

Foegen, A., & Marrison, C. (2010). Putting algebra progress monitoring into practice: Insights from the field. *Intervention in School and Clinic, 46.* Retrieved from http://search.proquest.com.proxy.library.vanderbilt.edu/

Foegen, A., Olson, J. R., & Impecoven-Lind, L. (2008). Developing progress monitoring measures for secondary mathematics: An illustration in algebra. *Assessment for Effective Intervention, 33*. doi:http://dx.doi.org/10.1177/1534508407313489

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. doi: http://dx.doi.org/10.1002/pits.20113

Ysseldyke, J., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003).

Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk*, *8*. doi: http://dx.doi.org/10.1207/S15327671ESPR0802_4