Using Curriculum-Based Measurement To Monitor Kindergarteners' Mathematics Development

Pamela M. Seethaler and Lynn S. Fuchs

Assessment for Effective Intervention 2011 36: 219 originally published online 12 July 2011
DOI: 10.1177/1534508411413566

Article

Hammill Institute on Disabilities

Assessment for Effective Intervention
36(4) 219–229

DOI: 10.1177/1534508411413566
http://aei.sagepub.com
SAGE

Using Curriculum-Based Measurement To Monitor Kindergarteners' Mathematics Development

Pamela M. Seethaler[1] and Lynn S. Fuchs[1]

Abstract

The purpose of this study was to examine technical and instructional features of a kindergarten curriculum-based measurement (CBM) tool designed to track students' mathematics progress in terms of computational concepts, procedures, and counting strategies. Students in 10 kindergarten classrooms in three elementary schools completed alternate forms of the CBM measure twice per month from January to May. Mathematics development was indexed on a standardized mathematics achievement test in May. Findings indicate strong reliability and validity of the CBM system, with coefficients exceeding .80 and .60, respectively. Technical features of the CBM system's skills analysis suggest implications for teachers' instructional decision-making.

Keywords

curriculum-based measurement, mathematics, elementary

Progress monitoring is an essential component of a response-to-intervention (RTI) framework for identifying students with mathematics difficulty (MD). Over the past several decades, a progress-monitoring technique known as curriculum-based measurement (CBM; Deno, 1985) has emerged as a reliable and valid method of gauging students' readiness for and success with school-based instruction (Foegen, Jiban, & Deno, 2007). Repeated and consistent sampling of overall competence in a given academic domain yields static information (i.e., students' score at a given point in time) as well as slope (i.e., growth on the measure over time). These data serve different purposes: Static information collected at a single point in time during the school year serves a screening purpose, whereas slope reflects students' response to instruction over time. Furthermore, a CBM system that features a curricular-sampling approach (Fuchs, 2004; Fuchs, Fuchs, & Zumeta, 2008) has the potential to provide a skills analysis for each student at varying points in the school year, highlighting students' relative strengths and weaknesses and assisting with teachers' instructional adaptations.

Studies show that deficient number combination knowledge (e.g., Jordan & Hanich, 2003) and poor use of counting strategies (e.g., Geary, Hoard, Byrd-Craven, & DeSoto, 2004) are hallmark manifestations of MD. These deficiencies should be identified early so they can be addressed quickly, in an attempt to offset future and more pervasive difficulty. Unfortunately, although single-skill screening assessments are available for identifying risk (e.g., Clarke & Shinn, 2004; Methe, Hintze, & Floyd, 2008), few assessments are available at the kindergarten level to monitor students' progress over time. The purpose of the present study was to evaluate the technical and instructional features of a kindergarten CBM measure designed to index overall competence by monitoring counting skill and number combinations knowledge over time. In this introduction, we briefly explain the stages of research necessary to validate progress-monitoring CBM. Then, we review prior work highlighting each stage of CBM research and finally explain how the present study extends previous work.

Stages of CBM Research

With respect to kindergarteners' risk status for MD, CBM is used to gauge students' response to instruction. Some researchers suggest that monitoring students' number sense, or informal knowledge of mathematical constructs and relations, may be the key to successful early identification and intervention (Berch, 2005; Geary, Bailey, & Hoard, 2009; Gersten, Jordan, & Flojo, 2005), as much as monitoring students' phonemic awareness and letter-sound knowledge predicts future reading difficulty (e.g., Schatschneider, Fletcher,

[1]Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Pamela M. Seethaler, 228 Peabody College, Vanderbilt University, Nashville, TN 37203, USA
Email: pamela.m.seethaler@vanderbilt.edu

Francis, Carlson, & Foorman, 2004). However, to validate the use of a CBM system as a means of screening students for potential MD, monitoring students' response to the curriculum, and informing teachers' instructional adaptation, it has been suggested that three stages of research must occur (Fuchs, 2004).

The first is to evaluate the technical features of the static score of a CBM probe administered at one point in time. Examples of the type of validation necessary at this first stage are internal consistency, alternate-form, or test–retest reliability, and concurrent or predictive validity as measured against some acknowledged standard. The second stage of research entails the evaluation of the technical features of the CBM slope. At this stage, alternate forms are administered to a group of students at multiple time points, and the data are analyzed to determine the slope of each student's progress; correlations between these slopes and important external criteria are studied. The purpose of this stage is to determine if increasing scores on the alternate CBM forms correspond to increasing competence with the construct under evaluation. The final stage of research necessary to validate the use of a CBM system focuses on the instructional usefulness of the resulting CBM data. At this stage, the purpose is to assess whether the data collected yield information that helps inform instruction for students and thus improve learning (e.g., Fuchs, Fuchs, Hamlett, & Stecker, 1991). Evidence from each stage of research is important for documenting separate but related features of utility for a particular mathematics CBM system. Although systematic evaluation at the kindergarten level has been increasing in recent years, support is not yet equally distributed across the three stages of research.

Summary of Previous Work

Foegen et al. (2007) provide a thorough review of progress-monitoring measures in mathematics for grades pre-K through eight. The majority of studies reviewed featured Stage 1 research. Limiting the pool to progress-monitoring research exclusively at the kindergarten level, we find a similar disproportionate focus on Stage 1 research. For example, the early numeracy measures developed by Clarke and Shinn (2004) have been repeatedly evaluated for evidence of their technical adequacy and predictive utility (e.g., Lembke & Foegen, 2009; Martinez, Missall, Graney, Aricak, & Clarke, 2009; Seethaler & Fuchs, 2010), but this focus is relevant to screening at one point in time. One of the early numeracy measures, Quantity Discrimination (QD), an individually administered, single-skill task of magnitude comparison, seems to show particular promise as a predictor of future MD risk with respect to Stage 1. Other research studies have focused on measures composed of single-skill tasks (e.g., Methe et al., 2008; VanDerHeyden, Witt, Naquin, & Noell, 2001). For example, VanDerHeyden et al. evaluated the reliability and validity of kindergarteners' ability to count a set of circles and choose the correct corresponding numerical amount, count a set of objects and write the corresponding amount, or draw a given amount of circles, whereas Methe et al. (2008) investigated the reliability and diagnostic utility of selected early numeracy skills in relation to end-of-year performance on established criterion measures. Across this small but growing body of literature, results support Stage 1 level of research, that is, with respect to the reliability and predictive utility of certain early numeracy CBM measures.

This reliance on Stage 1 research at the kindergarten level, however, results in an emphasis on screening students for risk and not on monitoring students for growth. Although this line of inquiry is useful, screening can be accomplished with measures other than those designed for progress monitoring. For example, school districts may use commercially available or district-created tests to measure knowledge of basic numeracy concepts. Furthermore, screening at the early grades for academic risk carries with it problems, most notably of false positives (i.e., erroneously identifying students in need of costly tutoring who would likely succeed in the absence of intervention; e.g., Fuchs et al., in press). The need to improve the accuracy of screening is critical in the area of mathematics, and a gated, two-stage screening process in which performance on static testing comprises the first stage, and progress monitoring or dynamic assessment represents the second (e.g., Compton, et al. 2010; Fuchs et al., in press), may emerge as a better option. At any rate, more research is needed at the second and third stages of CBM research, focusing on data from multiple time points, rather than remaining focused on the first stage.

With respect to Stage 2, some studies have examined the rate of growth of kindergarteners' mathematics performance across the year using slope as a predictor of mathematics outcome (e.g., Clarke, Baker, Smolkowski, & Chard, 2008; Jordan, Kaplan, Locuniak, & Ramineni, 2007). However, of the four early numeracy CBM measures administered in the Clarke et al. (2008) study (QD, oral counting, number identification, and missing number), only slope of performance on QD fit well enough to use in prediction models. One limiting factor may have been the nature of the measures: Single-skill tasks may not represent a robust indicator of overall mathematics achievement. Jordan et al. used a number sense core battery that featured items representing different kindergarten mathematics concepts. These authors found that slope of kindergarten mathematics performance, sampled four times across the entire kindergarten year and twice in first grade, accounted for 66% of the variance in first-grade mathematics outcome. This research provides evidence that growth over time on certain mathematics tasks may be linked to future mathematics development. However, for data to be used to judge response to instruction, frequent sampling of student performance is vital. The previous studies sampled behaviors twice (Clarke et al., 2008) and four times (Jordan et al., 2007) across the kindergarten year; future research is needed to evaluate rate of growth with more frequent data collection.

Studies that facilitate Stage 3 research in kindergarten mathematics are greatly needed and are few in the current literature. Fuchs et al. (1991) examined the role of CBM skills analysis in helping teachers develop instructionally sensitive adjustments for their students and effect better student outcomes. The authors randomly assigned teachers to receive CBM feedback comprising graphed scores only, graphed scores plus skills analysis, or no CBM (control). Results showed that teachers receiving feedback comprising graphed scores and skills analysis designed more responsive instructional changes and effected better student achievement than the competing groups. However, students in the Fuchs et al. (1991) study were in grades three through nine; to our knowledge, no work exists featuring this type of inquiry at the kindergarten level. A major goal of progress monitoring at kindergarten in mathematics is to help teachers tailor their instructional programs in ways that are responsive to the needs of students who are struggling with particular areas of early numeracy (Fuchs, 2004) and it is important to learn if the results of the Fuchs et al. (1991) study could be applied downward to younger students.

The skills analysis facilitated by a CBM system documents students' level of mastery of each of the component skills of the system. Investigating the extent to which the skills analysis promotes instructionally relevant changes (i.e., to improve student outcomes) represents Stage 3 research. Initially, however, the technical features of the skills analysis should be documented. That is, the skills analysis should yield reliable and valid information for teachers. Once that has been documented, the next step is to investigate its usefulness for enhancing instructional planning and student learning. At the present time, no CBM systems exist that provide such evidence for kindergarten mathematics.

Purpose of the Present Study

The purpose of the present study was to evaluate a kindergarten curricular sampling, multiple-skill mathematics CBM system sensitive to improvement in areas of mathematics known to be challenging for kindergarten children with MD, such as counting skill and number combinations acquisition. We assessed the technical features of the overall graphed score as well as the skills profile by comparing two consecutive skills analyses, the skills analysis from the beginning of March, and the skills analysis from the end of March. Stability of the skills analyses was determined by computing the percentage of agreement across the skills for the two time points for each student. The investigation of the technical properties of the skills analysis was modeled on earlier work conducted in mathematics at Grades 2 to 6 (Fuchs et al., 1994), in which these researchers coded each skill set within a skills analysis on a scale from 1 (*not tried*) to 5 (*mastered*).

Our research questions were as follows: What is the technical adequacy of the static, graphed scores of our kindergarten math CBM? What is the predictive validity of the CBM slope of improvement over seven testing occasions, spaced over 14 weeks of the kindergarten year? When we create a skills profile based on skills incorporated within the CBM kindergarten system, is the resulting skills profile reliable and valid?

Our kindergarten CBM probes are administered whole-class at the kindergarten level in a paper-and-pencil format. This type of testing format is not commonly used at kindergarten; most of the literature seems to favor individual testing situations. As such, we hypothesized that scores would be lower in the initial data collection events, not only because students would be unfamiliar with the testing format but also because they would not have progressed far enough through the curriculum to master the problem types incorporated within the multiple-skill CBM system. Furthermore, we expected students to perform near ceiling toward the end of the year on the kindergarten probes, as they neared mastery of the problem types. Thus, we expected that slope might be relatively flat for students who start the program with higher mathematics skills and who do not have as far to grow as their lower-achieving peers. With respect to technical adequacy of the static scores, we hypothesized that levels of internal consistency and test–retest reliability would be above .80 and that measures of concurrent and predictive validity would range from .49 to .74, given previous work investigating this measure with a different sample of kindergarten students (Seethaler & Fuchs, 2010). With respect to the predictive validity of measures of initial CBM score and of slope of growth across time, we expected slope to correlate less well with mathematics development, given that some high-achieving students may show little to no growth on the measure because of ceiling effects. With respect to the skills profile, because students would be taking the tests frequently (i.e., approximately every other week), we expected skills profiles to remain relatively stable across two data points.

Method

Participants

We randomly selected 10 kindergarten teachers from three schools in a southeastern metropolitan school district from a pool of 18 teachers interested in participating. Six classrooms were from schools with Title I funding (i.e., a high percentage of the students were from low-income families). Teachers reported their own demographic information. At the time of the study, four teachers were 21 to 29 years old, one was 30 to 39 years old, one was 40 to 49 years old, three were 50 to 59 years old, and one was older than 60 years. The majority of the teachers were female (90%) and African American (80%; two were White). Teachers reported their highest level of education as a bachelor's degree (50%), master's degree (30%), or master's degree plus 30 credit hours (20%) and had been teaching an average of 14.4 years. Teachers reported their class size as a mean of 18 students, with fewer than 1 student per class (0.6) receiving special education services.

This particular school district has a transient population, with students transferring in and out of schools quite regularly. Teachers consented to adopt the progress-monitoring measures as part of their classroom mathematics program. Altogether, 193 students completed at least one of the seven progress-monitoring tests. However, some students remained for a short time (in some cases, less than 1 month) before exiting the class. Thus, we excluded students for whom fewer than four data points were available, leaving a sample of 180 students.

From this larger sample of students, we obtained parental consent and student assent from 87 students across the 10 classrooms for additional, individual testing. Teachers reported demographic information for this subset of group of students. The average age of students at the time of end-of-kindergarten testing was 6.0 years ($SD = 0.33$). Of the students in this group, 50% were male and 60% received free or reduced-price lunch. The racial demographics of the participants closely resembled that of the district from which they were sampled: 55% of the study sample was African American, 30% White, 12% Hispanic, and 3% other, compared to 47%, 33%, 16%, and 4%, respectively, of the district population. Seven percent of the students received special education, all for speech disability. Approximately 2% of the students qualified as English learners, and 2% had been retained for 1 year. Teachers reported that approximately 28% of the students were above grade level in mathematics skill, 55% were at grade level, and 17% were below grade level.

Measures

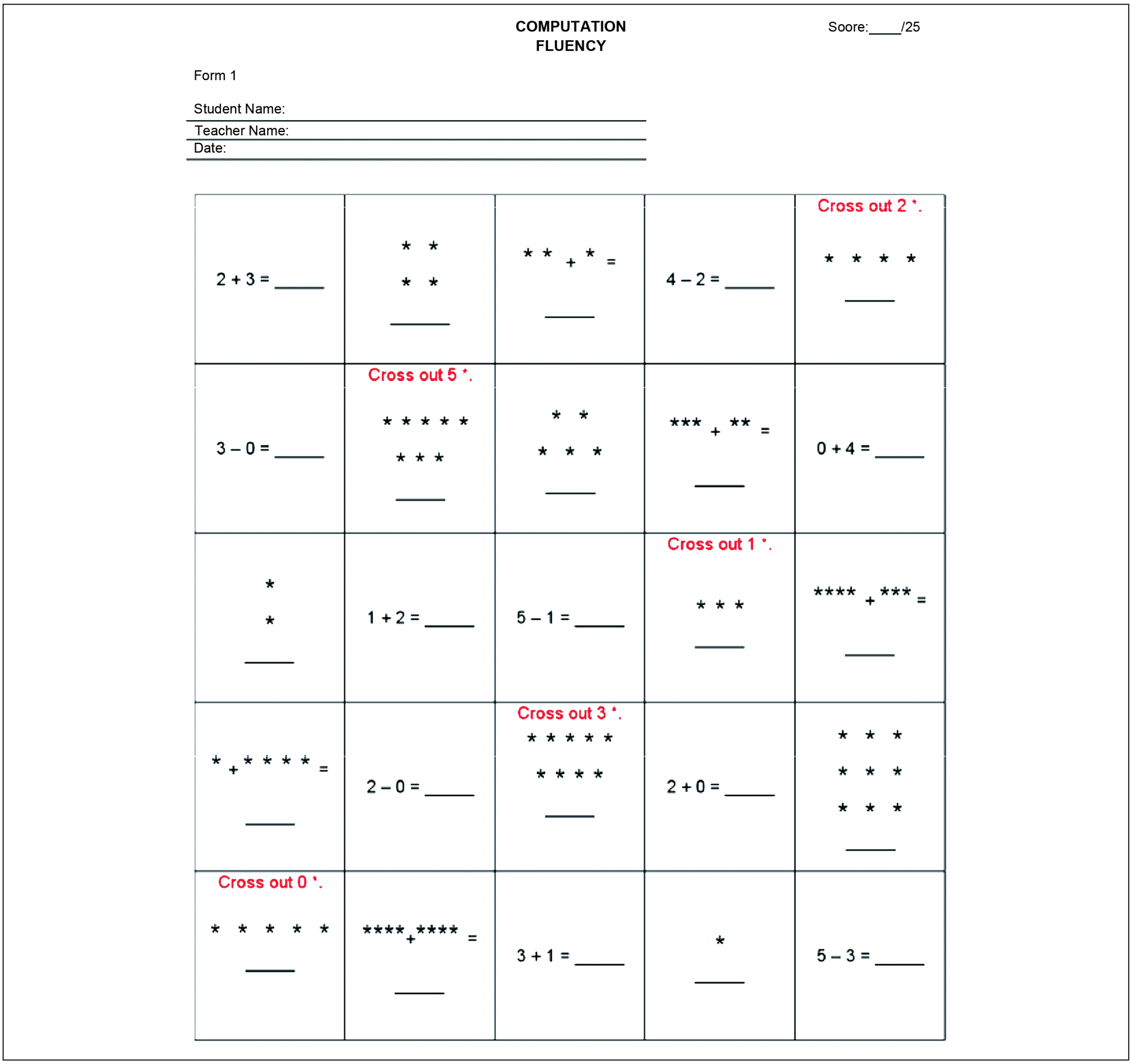
CBM progress monitoring. The measure used for progress monitoring, Computation Fluency (CF; Seethaler & Fuchs, 2010), is a 5-min timed assessment of counting, addition, and subtraction fluency. It is administered in a whole-class setting and includes 25 items (5 items each of five problem types) presented in random order on one side of an 8.5- by 11-in piece of paper. The five types of problems are counting stars in a single set (Counting Stars), counting two sets of stars (Adding Stars), subtracting crossed-out stars from a single set (Subtracting Stars), addition facts with numerals (presented without star icons; Addition Facts), and subtraction facts with numerals (presented without star icons; Subtraction Facts). The measure has five rows of five problems each, bordered in dark lines to help delineate items from one another. The student is not penalized for number reversals or poorly formed written responses. Performance is scored as number of correct items in the given time. In addition, each item on each CF form was coded for type of skill (i.e., counting stars, counting sets of stars, subtracting sets of stars, addition facts, subtraction facts) so that percentage correct of type of skill was available for each form.

CBM probe construction. We created 20 forms, identical in format but differing in actual items; we used the first seven forms for the present study. CF is conceptually based on the CBM probes for Grades 1 through 6 developed by Fuchs and colleagues (e.g., Fuchs et al., 1994; Fuchs, Fuchs, & Compton, 2004). It resembles the Computation CBM probes and relies on a curricular-sampling approach, as do the CBM probes. For analyses, we used an average of the first two CF scores to quantify beginning level of mathematics ability, an average of the last two CF scores for ending level. See Figure 1 for an example of CF.

End-of-year outcome and definition of difficulty. We measured kindergarteners' mathematics skill at the end of the year with the Test of Early Mathematics Ability–3rd Ed. (TEMA; Ginsburg & Baroody, 2003). The TEMA's two forms (Form A and Form B) assess early mathematics skill from the following domains: numbering skills, number-comparison facility, numeral literacy, mastery of number facts, calculation skills, and understanding of concepts. We used Form A. There are 72 items of increasing difficulty. The examiner presents the student with several trials for each item; the student must pass a specified number of trials (e.g., two trials correct out of three) to earn 1 point for the item. Testing is discontinued after five consecutively missed items. Performance is noted as number of correctly answered items; the session is not timed. According to the manual, coefficient alpha for 6-year-olds for Form A is .95. Students received a designation of MD if they scored below the 16th percentile on the TEMA.

Procedure

Principals of schools were contacted in the beginning of January. We then met with kindergarten teachers from interested schools to explain the nature of the study and to answer questions. Of the 18 teachers who volunteered to participate, we randomly selected 10. Teachers sent home consent forms to parents to elicit permission for their children's participation in the individual testing potion of the study. Approximately 48% of the forms were returned. Research staff, comprising graduate students with varying degrees of classroom experience, were trained by the first author in mock administrations to 100% accuracy, using a checklist of administration directions. Research staff administered the first whole-class progress-monitoring measure to students during the last week in January, following a 15-min whole-class lesson comprising a practice session with sample items. Teachers then met with the first author and were trained to 100% accuracy, using a checklist of administration directions, for administering CF. After the initial administration of CF in January by research staff, teachers administered an alternate form to their class twice per month, on days specified by the first author. Teachers read from scripts and used timers to ensure consistency of testing administration across probes. They collected all tests and returned them to research staff, who scored each test via a computer scoring program, entering the data into a database. All of the tests were scored and entered into a second database a second time by an independent scorer and the two databases were compared for discrepancies. All discrepancies were resolved by examining



COMPUTATION FLUENCY Soore:____/25

Form 1

Student Name:

Teacher Name:

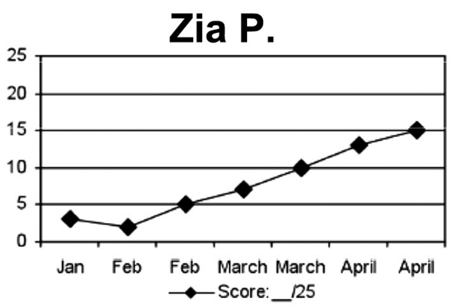
Date:

2 + 3 = _____	* * * * ______	* * + * = _____	4 – 2 = _____	Cross out 2 *. * * * * _____
3 – 0 = _____	Cross out 5 *. * * * * * * * * _____	* * * * * _____	*** + ** = _____	0 + 4 = _____
* * _____	1 + 2 = _____	5 – 1 = _____	Cross out 1 *. * * * _____	**** + *** = _____
* + * * * * = _____	2 – 0 = _____	Cross out 3 *. * * * * * * * * * _____	2 + 0 = _____	* * * * * * * * * _____
Cross out 0 *. * * * * * _____	****+**** = _____	3 + 1 = _____	* _____	5 – 3 = _____

Figure 1. Example of computation fluency kindergarten curriculum-based measurement form

the original protocols. In this way, we ended up with a single database, free from data-entry error. Once per month, during the 1st week of the month, research staff delivered a teacher report to each teacher. The report included a letter explaining the results; a CBM graph for each student, showing each student's progress across all data points collected to date; and an aggregated data page, listing each student's overall score as well as each student's percentage correct across each type of item on the test. See Figure 2 for an example of a student graph and whole-class data as reported to teachers.

One week after the last CF test was administered, research staff administered the TEMA to students who were participating in the individual testing portion of the study. Testing took place either in a quiet hallway outside of the classroom or in the library. Students were given a small prize (i.e., pencil or toy) at the end of the session. Research staff scored the tests and entered the data into a computer scoring program. All the tests (100%) were rescored and the data were entered a second time; all discrepancies were resolved by examining the original protocols. In addition, all individual testing sessions were audiotaped via digital recorder and 18% of the sessions were evaluated by the first author for fidelity of testing administration. A detailed checklist was used to ensure each test was administered and scored by the tester as it was intended and



April Teacher Report: Scores from Mar CBM Tests 4 and 5

	Score: __/25	% Correct: Counting Stars	% Correct: Adding Stars	% Correct: Subtracting Stars	% Correct: Addition Facts	% Correct: Subtraction Facts
Student	7.5	0%	0%	50%	40%	60%
Student	25	100%	100%	100%	100%	100%
Student	10.5	60%	60%	10%	50%	30%
Student	23	100%	100%	80%	100%	80%
Student	23	100%	90%	100%	90%	80%
Student	22	90%	80%	90%	90%	90%
Student	10	100%	0%	10%	50%	40%
Student	24	100%	100%	100%	90%	90%
Student	11	80%	80%	20%	20%	20%
Student	23	100%	90%	100%	100%	70%
Student	7	30%	30%	20%	40%	20%
Student	14	80%	30%	80%	50%	40%
Student	7.5	70%	30%	10%	10%	30%
CLASS AVERAGE:	16.0	78%	61%	59%	64%	58%

Figure 2. Example of student graph and class data included in the kindergarten CBM system's monthly report

described in the manuals by each test developer. The digital recorder was placed in close proximity on the testing table and captured all tester and student oral interactions; original test protocols were examined to ensure written products were scored accurately. Interscorer agreement (the number of agreed points divided by the total number of points) was 98%.

Results

Table 1 includes means, standard deviations, and reliability coefficients. We report data separately for the larger sample of students for whom we have at least four CBM data points (n = 180) and the subset of students for whom we have at

Table 1. Means and *SD* for CF[a] and End-of-Kindergarten Mathematics Measures

	n = 180	*n* = 87
Measures	*M* (*SD*)	*M* (*SD*)
CF initial[b]	12.15 (5.72)	12.94 (6.34)
CF final[c]	17.90 (6.41)	18.45 (6.21)
Slope[d]	0.54 (0.48)	0.53 (0.49)
SEE[e]	2.44 (1.18)	2.29 (1.19)
TEMA[f] (raw score)		32.76 (9.09)
TEMA (standard score)		99.91 (9.09)

a. CF is computation fluency.
b. CF Initial is average of first two CF tests.
c. CF Final is average of last two CF tests.
d. Slope is rate of growth as computed with least squares regression between biweekly testing occasions and CF scores.
e. SEE is standard error of estimate from regression.
f. TEMA is Test of Early Mathematics Ability, 3rd Ed. (Ginsburg & Baroody, 2003).

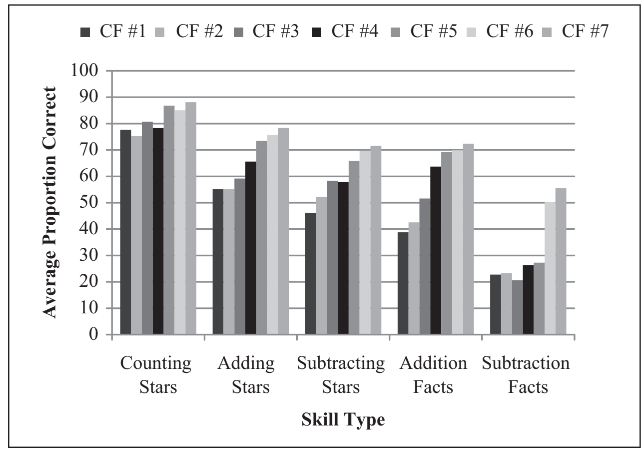


Figure 3. Proportion correct of each of five skill types, averaged across administrations one through seven of computation fluency (CF)

least four CBM data points and TEMA data (n = 87). We used the average score on the first two CF administrations (i.e., last week of January and 2nd week of February) to represent initial mathematics level; the average score of the last two administrations (i.e., 2nd and last week of April) to represent final level. Slope was computed with least squares regression between biweekly testing occasions and CF scores (and then divided by 2 to derive the weekly rate of increase); SEE is the standard error of the estimate. Slope calculated this way was the same as when calculated with hierarchical linear modeling; a negligible amount of variance (i.e., 3%) was attributed to the classroom level.

Reliability and Validity

We assessed stability of the graphed scores by computing correlations between each administration and the previous administration. Stability ranged from .80 to .87.

Following Fuchs et al. (1994), we assessed the stability of the skills analysis by comparing the relation between two consecutive skills analyses from the two CF forms administered in March (i.e., CFs 4 and 5). First, we determined the percentage correct of each skill type for each student. With five items of each skill type, the possible percentages correct for students were 0, 20, 40, 60, 80, or 100. Figure 3 shows the proportion correct of each of the five skill types, averaged across the seven administrations. The mean percentage correct across all forms was 81.14 (SD = 20.57) for Counting Stars, 65.71 (SD = 28.50) for Adding Stars, 59.70 (SD = 33.05) for Subtracting Stars, 58.12 (SD = 31.42) for Addition Facts, and 32.26 (SD = 222.51) for Subtraction Facts. We then compared performance on CF 4 with that on CF 5 by subtracting the smaller number by the larger number, subtracting that difference from 100, and dividing that difference by 100 to yield percentage correct. (For example, if a student scored 60 for proportion correct on Adding Stars items on CF 4 and 80 for the same skill type on CF 5, we subtracted 60 from 80, then subtracted 20 from 100, then divided 80 by 100 for 80%.) After calculating the percentage of agreement for each student for each skill type, we averaged the percentages for skill type. For Counting Stars, the percentage of agreement between CFs 4 and 5 was 85.84 (SD = 22.04); for Adding Stars, Subtracting Stars, Addition Facts, and Subtraction Facts, 81.04 (SD = 23.74), 85.45 (SD = 20.23), 84.68 (SD = 18.55), and 87.14 (SD = 17.66), respectively.

With respect to concurrent and predictive relation of the CBM graphed (static) scores (i.e., initial and final performance on the CF forms) and the slope of growth of development across the 14 weeks, we calculated correlations between the CBM data (static scores and slope) and the TEMA (administered at the end of kindergarten): .61 (p < .001) between initial CBM score and TEMA and .69 (p < .001) between final CBM score and TEMA. To compute the correlation between slope and TEMA, we considered the subset of students whom teachers would monitor in actual practice: at risk who fail the universal screen (see predictive utility analysis description in next section). For these children, we calculated the correlation between slope and end-of-year mathematics skill: .49 (p < .001). This is similar to Fuchs et al. (2004), who evaluated the predictive validity of first-grade CBM measures with respect to end-of-year reading performance. They found correlations from .27 to .54 for fall administration and from .32 to .63 for spring.

Following Fuchs et al. (1994), we indexed the validity of the skills analysis by correlating a composite score from the skills profile of the second CBM administration of each month (excluding the 1st month, in which students were administered only one CBM) with the students' averaged graphed scores from the same month. For example, the composite score from the skills profile based on CF 7, at the end of April, was correlated with the averaged graphed scores from CFs 6 and

7 administered in the same month. The correlation for the 2nd, 3rd, and 4th month was .95, .96, and .96, respectively.

Predictive Utility of Initial CBM Performance Used for Universal Screening

We used logistic regression and area under the receiver operating characteristics (ROC) curve (AUC) to evaluate the predictive utility of initial performance on the kindergarten CBM measures for classifying MD risk at the end of kindergarten. Sensitivity, which is the proportion of children in the sample predicted by the model to be MD (in this study, 20), is computed by dividing the number of true positives by the sum of true positives and false negatives. Specificity, which is the proportion of students correctly predicted to not have MD, is computed by dividing the number of true negatives by the sum of true negatives and false positives. AUC, a measure of discrimination (McClish, 1989), is a plot of true-positive rate against the false-positive rate for the different possible cutpoints of a test. AUC ranges from .50 (chance) to 1.00 (perfect prediction); AUC below .70 indicates a poor predictive model, .70 to .80 indicates a fair model, and .80 to .90 or .90 to 1.00, a good or excellent model (e.g., Fuchs et al., 2007). Results of the logistic regression analysis were as follows. Holding sensitivity to 90.0% to minimize false negatives, initial CBM score predicted end-of-year MD status with specificity of 63.6% and hit rate of 69.8%. AUC was .77 (confidence interval of .66–.87), which is deemed fair. This model resulted in 2 false negatives and 24 false positives.

Discussion

The purpose of the present study was to contribute to Stage 1 and Stage 2 research on a kindergarten curricular-sampling, multiple-skill mathematics CBM system, even as we paved the way for Stage 3 research by investigating technical features of the skills analysis. To assess the technical adequacy of CBM static scores, a requisite first step in investigating the validity of a CBM system, we evaluated the test–retest reliability and concurrent and predictive validity of students' initial and final performance on alternate CBM forms. Findings indicate strong reliability, with coefficients exceeding .80. This corroborates previous findings for this particular measure (Seethaler & Fuchs, 2010) as well as other research on other curricular-sampling CBM mathematics systems (e.g., Foegen et al., 2007). We further compared kindergarten students' static performance on initial CBM with end-of-year performance on a standardized, global measure of mathematics, documenting predictive validity of .61. Comparing final CBM performance with the same global measure, concurrent validity was .69. This is in line with previous work for this kindergarten curricular-sampling system (Seethaler & Fuchs, 2010), which documented coefficients between .49 and .74. Present findings together with Seethaler and Fuchs support Gersten et al. (2005), who suggested early numeracy CBM measures may be reliable and valid predictors of potential MD in kindergarteners.

Of course, computing predictive validity correlations is not the same as evaluating a CBM test's accuracy in predicting students' risk status for MD. Practically speaking, accurate prediction is the goal when using CBM as a universal screener; teachers require accuracy so that they can identify students for early intervention. Toward that end, we used logistic regression and AUC analyses to investigate how well initial performance on our CBM measure would classify students as MD, based on low performance on the TEMA, a global measure of mathematics ability, administered at the end of kindergarten. By setting the cutpoint at the score that limited false negatives, initial CBM resulted in high sensitivity (i.e., 90.0%) but low specificity (i.e., 63.6%) with a large number of false positives. This is problematic because excessive false positives stress time, effort, and financial resources of a school. Excessive false positives have been documented in reading as well when a one-stage universal screen is employed (e.g., Johnson, Jenkins, Petscher, & Catts, 2009). This underscores the need to develop a two-stage process of screening to enhance classification accuracy (Fuchs et al., in press).

Such a second stage of gated screening may entail progress monitoring. Once Stage 1 CBM research have been satisfied, the next step concerns evaluating students' growth over time on alternate forms of the system (Fuchs, 2004). The curricular-sampling approach we used to construct our kindergarten mathematics CBM led us to sample items students are expected to master by the end of the year. As such, we hypothesized that initial scores (sampled half-way through the year) would be lower than scores sampled at the end of the year, after students had experienced greater exposure to mathematics instruction. Our results confirmed this. As Table 2 shows, average scores increased incrementally from the first through the seventh administration. This increase represents an average slope (or rate of improvement over time) of 0.54 problems correct ($SD = 0.48$) per week for the larger sample of students. (Note: Similar results were found for the subset of students.) The steady increase across the 14 weeks of assessment suggests that our measure was sensitive to growth in counting skill and arithmetic number combinations, the underlying constructs of our measure.

Information about how well students are improving on these core skills may benefit teachers' instructional planning. For example, a student with a flat slope indicates inadequate mathematics learning, warranting prompt intervention; by contrast, a positive slope indicates satisfactory response to instruction. Few studies have documented evidence supporting Stage 2 research for kindergarten mathematics CBM measures (see, e.g., Clarke et al., 2008; Jordan et al., 2007). Future work should continue investigating the technical features of slope, a unique contribution of progress-monitoring systems and an important indicator of students' response to classroom instruction.

Table 2. Means, *SD*, and Reliability of Computation Fluency Forms

Computation Fluency (CF) Form	*M* (*SD*)	Alpha	Test–Retest[a]
CF 1	12.02 (5.71)	.88	
CF 2	12.42 (6.13)	.89	.83
CF 3	13.52 (6.28)	.91	.80
CF 4	14.59 (6.80)	.93	.81
CF 5	16.13 (6.08)	.92	.85
CF 6	17.52 (6.81)	.94	.82
CF 7	18.29 (6.37)	.93	.87

a. Test–retest is the correlation between a CF form and the previously administered form; approximately 2 weeks elapsed between administrations.

Implications for Practice

Stage 3 research is arguably the most essential aspect of formative evaluation. Assessment for effective intervention represents a major goal of kindergarten mathematics CBM, particularly with respect to RTI. In the absence of a data-driven system for documenting students' success and failures with classroom instruction, teachers may not accurately and sensitively design academic intervention to meet students' needs. Stage 3 CBM research, according to Fuchs (2004), is when studies are conducted to investigate if teachers can use the data from CBM to inform instructional decisions and, importantly, improve student achievement. As a prerequisite, research must document that the CBM system provides instructionally relevant and technically adequate information that can be linked directly to the curriculum. In the present study, we did not specifically evaluate teachers' instructional adaptations as a function of CBM results. Instead, we addressed the prerequisite steps for conducting Stage 3 research with our kindergarten mathematics CBM system by examining the technical features of the skills analysis it provides.

We sampled kindergarten skills emphasizing counting skill and arithmetic number combinations, hallmark areas of deficit for students with MD in the primary grades (Geary et al., 2004; Jordan & Hanich, 2003). The five skills comprising the test ranged in difficulty from counting a set of star icons and writing the corresponding digit (i.e., Counting Stars) to subtraction of number combinations presented without icons (i.e., Subtraction Facts). Interestingly, the skill types all showed a trend toward growth over time (see Figure 3) and the skill types retained their relative order of difficulty across the 14 weeks. That is, Counting Stars remained the easiest type of problem; Subtraction Facts, the most difficult. Adding Stars, Subtracting Stars, and Addition Facts appeared more variable in their relative difficulty across the weeks. Future research should investigate if strengths or weaknesses in one area contribute more variance in predicting future mathematics skill; such information would be helpful in designing more accurate screening systems.

We disaggregated results from each form by skill type to assess the reliability and validity of the skills profiles, in similar fashion to earlier work (Fuchs et al., 1994). Results showed that the skills profiles were stable across forms (percentage of agreement between 81.04 and 87.14), even as students showed growth over time. Furthermore, information summarized across the skills analysis and the graphed scores correlated strongly (between .95 and .96), suggesting that the information from the skills profile represents a valid representation of the overall information provided by the graphed score.

Findings are important for several reasons. First, we replicated and corroborated work on static CBM scores, lending further support to the technical adequacy of kindergarten mathematics CBM. Second, stability of students' scores suggests the feasibility of administering a brief, timed, whole-class measure to kindergarten students. This is important in that most literature to date assesses technical features of individually administered kindergarten CBM systems, which can be time consuming and less practical for practitioners with large numbers of students. Teachers may be more likely to collect CBM data frequently with group administration, because of ease of administration, providing more frequent observations of students' growth. This, of course, needs to be tested empirically. Third, we extended the existing body of kindergarten mathematics CBM research by providing evidence of the technical features of a skills analysis. By doing so, we set the stage to begin research on the instructional utility of the skills profiles with a randomized control trial at the kindergarten level, much as Fuchs et al. (1991) did at Grades 3 through 9. Clearly, it is important to document the technical adequacy of CBM static scores and slope, ensuring that resulting scores are reliable and obtained slopes are meaningful. However, if teachers are expected to adjust instruction for students who are struggling with basic numeracy and are most at risk for developing MD, we must provide them with tools that not only indicate the students' current level of functioning and response to instruction but also guide their instructional decisions.

As readers interpret findings, however, they should consider the following limitations. First, participants were selected from one school district in one metropolitan location. Future research should sample students representing a more diverse population to allow for greater generalizability of results. Second, we did not include norm-referenced measures of overall mathematics skill level, so we do not know the extent to which this may influence the predictive utility of the progress-monitoring system. In future work, such measures should be included. Finally, it should be noted that teachers received monthly feedback of their students' mathematics progress (or lack thereof), yet we did not monitor the extent to which teachers adjusted their instruction in response to these data. A focus on the potential of teacher effects represents the important next step in the evaluation of our kindergarten mathematics CBM system.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by Award R324A090039 from the U.S. Department of Education and by Award R01HD059179 and Core Grant HD15052 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to Vanderbilt University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. Department of Education or the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

References

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38*, 333–339.

Clarke, B., Baker, S., Smolkowski, K., & Chard, D. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education, 29*, 46–57.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*, 234–248.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., . . . Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102*, 327–340.

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

Foegen, A., Jiban, C., & Deno, S. L. (2007). Progress monitoring measures in mathematics: A review of the literature. *Journal of Special Education, 41*, 121–139.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement. *School Psychology Review, 33*, 188–192.

Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Hamlett, C. L., & Seethaler, P. M. (in press). Two-stage screening for math problem-solving difficulty using dynamic assessment of algebraic learning. *Journal of Learning Disabilities*.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*, 7–21.

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring in first grade: Implications for responsiveness to intervention. *Exceptional Children, 73*, 311-330.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). The role of skills analysis in curriculum-based measurement in math. *School Psychology Review, 19*, 6–22.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Thompson, A., Roberts, P. H., Kubec, P., & Stecker, P. M. (1994). Technical features of a mathematics concepts and applications curriculum-based measurement system. *Diagnostique, 19*, 23–49.

Fuchs, L. S., Fuchs, D., & Zumeta, R. O. (2008). A curricular-sampling approach to progress monitoring: Mathematics concepts and applications. *Assessment for Effective Intervention, 33*, 225–233.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool. *Journal of Psychoeducational Assessment, 27*, 265–279.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with arithmetical disability. *Journal of Experimental Child Psychology, 88*, 121–151.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*, 293–304.

Ginsburg, H., & Baroody, A. (2003). *Test of Early Mathematics Ability* (3rd ed.). Austin, TX: PRO-ED.

Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice, 24*, 174–185.

Jordan, N. C., & Hanich, L. B. (2003). Characteristics of children with moderate mathematics deficiencies: A longitudinal perspective. *Learning Disabilities Research & Practice, 18*, 213–221.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice, 22*, 36–46.

Lembke, E., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learning Disabilities Research & Practice, 24*, 12–20.

Martinez, R. S., Missall, K. N., Graney, S. B., Aricak, O. T., & Clarke, B. (2009). Technical adequacy of early numeracy curriculum-based measurement in kindergarten. *Assessment for Effective Intervention, 34*, 116–125.

McClish, D. K. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making, 9*, 190–195.

Methe, S. A., Hintze, J. M., & Floyd, R. G. (2008). Validation and decision accuracy of early numeracy skill indicators. *School Psychology Review, 37*, 359–373.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparison analysis. *Journal of Educational Psychology, 96*, 265–282.

Seethaler, P. M., & Fuchs, L. S. (2010). The predictive utility of kindergarten screening for math difficulty. *Exceptional Children, 77*, 37–59.

VanDerHeyden, A. M., Witt, J. C., Naquin, G., & Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review, 30*, 363–382.

About the Authors

Pamela M. Seethaler, PhD, is a research associate in the Department of Special Education at Vanderbilt University. Her research interests include the early identification of mathematics disability and the development of effective practices for mathematics instruction for students with learning disabilities in the primary grades.

Lynn S. Fuchs, PhD, is Nicholas Hobbs Professor of Special Education and Human Development at Vanderbilt University. She conducts research on instructional and assessment issues pertaining to reading and mathematics learning disabilities.