Progress Monitoring Measures in Mathematics

A Review of the Literature

Anne Foegen
Iowa State University
Cynthia Jiban and Stanley Deno
University of Minnesota

This review of literature on progress monitoring was designed to examine the full array of curriculum-based measures (CBMs) in mathematics for students from preschool to secondary schools. We organized the article around two primary concerns: the approach used to develop the measures (curriculum sampling or robust indicators) and the type of research necessary to establish the viability of the tasks. Our review addressed the technical adequacy of the measures as indicators of performance and progress, as well as teachers' use of the measures to improve achievement. The largest number of studies has been conducted at the elementary level, with less work in early mathematics or at the secondary level. In general, the measures have acceptable levels of reliability; the criterion validity of mathematics CBMs appears to be lower than that for reading CBMs. One important finding that emerged was the relatively low degree of consensus on the best approach to use in developing mathematics CBMs. We discuss probable reasons for this, along with implications for practice and research.

Continuous monitoring of individual student academic progress has long been an important aim within the field of special education (Deno, 1985; Fuchs, 2004). Curriculum-based measurement (CBM) represents an empirically supported system of progress monitoring that has produced demonstrated effects on student achievement, particularly in reading (Fuchs, Deno, & Mirkin, 1984; Fuchs, Fuchs, & Hamlett, 1989b; Jones & Krause, 1988). In recent years, as general educators and policymakers have emphasized greater accountability for schools' efforts to teach all children, many researchers and practitioners have asserted the utility of progress monitoring—particularly in reading—for increasing numbers of students, regardless of disability status (Deno, 2003). Given growing attention to student achievement in mathematics, the practice of continuous progress monitoring should be similarly scaled up in this content domain. To do so, education professionals need to take stock of the empirical evidence for mathematics progress monitoring measures with regard to technical adequacy and instructional utility.

The purpose of this article is to review the existing empirical literature on mathematics progress monitoring measures. As we began our work on the mathematics strand of the Research Institute on Progress Monitoring (RIPM), we sought to identify the areas of greatest need in which to focus our research. A review of the literature provided the best means of evaluating the current status of research in mathematics CBMs. We have organized our article using two concerns identified by Fuchs (2004): (a) strategies for developing CBM tasks and (b) stages of research necessary to establish the viability of those tasks.

Two Approaches for Developing CBMs in Mathematics

Fuchs (2004) described two broad approaches for the development of CBM tasks. This categorization is particularly useful in the area of mathematics. In one approach, termed *curriculum sampling*, researchers have developed measures by constructing representative samples of the year's mathematics curriculum—taking at second grade, for instance, a larger proportion of addition and subtraction problems and, at sixth grade, a sampling that includes more advanced skills, such as division of decimals or addition of fractions. The curriculum sampling approach has been applied to computation, as well as to conceptual problems and applied mathematics skills. For the second approach, termed *robust indicators,* researchers have sought to identify measures that represent broadly defined proficiency in mathematics. Using this approach, effective measures are not necessarily representative of a particular curriculum, but are instead characterized by the relative strength of their correlations to various overall mathematics proficiency criteria. Measures in this second approach attempt to parallel in mathematics the kind of "robustness" that the oral reading CBM task offers in the area of reading: not necessarily drawn from the student's yearly curriculum, yet offering

Address: Anne Foegen, Curriculum and Instruction, Iowa State University, N162D Lagomarcino Hall, Ames, Iowa 50011-3198.
E-mail: afoegen@iastate.edu

strong correlations to a host of criterion measures of overall subject area proficiency. In the following section, we consider the relative merits of each approach and their implications for the development of mathematics CBMs. Practitioners will need to weigh these advantages and disadvantages when making decisions about which mathematics measures to use for progress monitoring.

A primary advantage of the curriculum sampling approach is the direct link it provides to the instructional curriculum, which facilitates the means to provide teachers with diagnostic feedback about a student's performance with regard to specific skills or concepts. These data may assist teachers in designing effective remedial instruction. This direct link to the curriculum also engenders limitations associated with this approach. Given the high level of curriculum specificity in mathematics, the curriculum sampling approach requires that different measures be developed to mirror each year's mathematics curriculum. As a result, the measures model student growth only within a single year, not across multiple years of learning. Moreover, current curriculum programs implemented in the nation's schools are quite diverse, with little consensus on placement of concepts in an instructional sequence or relative emphasis among topics (Reys, Dingman, Sutter, & Teuscher, 2005). The curriculum sampling approach may result in measures that are wedded to a particular curriculum program, necessitating the development of multiple CBM systems, each linked to a specific mathematics curriculum.

The robust indicators method offers several advantages from RIPM's perspective. As the RIPM research team strives to create a seamless and flexible system of progress monitoring measures in mathematics, we are particularly interested in measures that can be used across multiple grade levels; robust indicators hold the promise of doing this because they are not linked to any particular curriculum model or instructional sequence. Instead, the search for robust indicators represents an effort to identify aspects of core competence in mathematics. This measurement of core competence is designed to produce robust indicator data that are predictive of important outcomes in mathematics, regardless of the vagaries of specific curriculum programs or high stakes state tests. Moreover, the use of robust indicators enables students' growth to be modeled over multiple years of learning. Because the robust indicators tap core aspects of mathematics performance, rather than a broad range of skills and concepts from the instructional curriculum, they are less useful to teachers in providing diagnostic information about students' strengths and weaknesses in particular areas.

A Continuum of Research on Curriculum-Based Measures

A three-stage continuum of CBM research described by Fuchs (2004) is particularly applicable to work in mathematics. As measures are initially developed, Stage 1 research involves exploring the technical adequacy of their use as static (one point in time) indicators. This research emphasizes the reliability and criterion validity of the measures and the extent to which scores can be used to predict future performance or achievement of important outcomes. Stage 2 research examines the technical characteristics of slopes generated when the measures are used for continuous progress monitoring. Research at this stage examines the variability in repeated measurement and the extent to which slopes are reflective of student growth in the content area. Stage 3 research includes studies examining the instructional utility of the measures. These applied studies investigate whether teachers' use of the measures to inform their instructional decisions results in improved student achievement.

Our review begins by describing the procedures we used to identify studies of CBM mathematics measures. We then provide a general overview of the body of Stage 1 and 2 studies, noting approaches used to provide evidence of technical adequacy and suggesting guidelines for readers in evaluating the evidence provided for each type of measure. Following a summary of the technical adequacy evidence, we address the existing research on using mathematics CBMs to improve student achievement (Stage 3). We conclude the article with a discussion of the current state of empirical evidence for CBMs in mathematics and offer recommendations for both practitioners and researchers.

Method

The goal of RIPM is to create a seamless and flexible system of progress monitoring measures in reading, written expression, and mathematics. More specifically, we are particularly interested in identifying or developing measures suitable for students of diverse ability levels across multiple grades. Our research team conducted a comprehensive review of the CBM literature in reading, written expression, and mathematics. The goal was to identify gaps in the technical adequacy evidence in each content area to best focus our research efforts.

Research team members searched electronic databases including ERIC, Science Citation Index Expanded, PsycInfo, Digital Dissertation, and the Expanded Academic Index using the following terms: *curriculum based measurement, curriculum-based measurement, curriculum based measure, curriculum-based measure, general outcome measure*, and *progress monitoring*. This process yielded 578 articles, dissertations, and reports related to CBM. Titles and abstracts were screened to confirm that they were related to CBM, and Method sections were screened to identify those that reported results of empirical studies of CBM (i.e., included CBM as a dependent or an independent variable), yielding 160 articles. In addition, the complete set of technical reports produced by the Institute for Research on Learning Disabilities (IRLD) at the University of Minnesota was accessed and included in the review. These documents were then reviewed by a team of educational psychology graduate students, grouped by subject (reading, mathematics, spelling, and writing), and coded for study characteristics, such as sample size and demographics, type of CBM and criterion measures used, and administration and scoring procedures. Twenty-nine (18%) of the articles in this initial search addressed mathematics measures. We added additional papers as they were published. Our focus in this article is limited to published studies and IRLD reports that address Fuchs's (2004) three stages of research on mathematics CBMs. The one exception to our inclusion rules was a conference paper presented by Espin, Deno, Maruyama, and Cohen (1989) that reports results for their *Basic Academic Skills Sample* (BASS), an instrument that has been subsequently used in many research projects.

This process yielded a total of 32 reports on mathematics progress monitoring measures. We summarize the studies that address the technical adequacy of the measures (research representing Stages 1 and 2) in Table 1 according to the sample characteristics, the type of measure, and the results reported for reliability, criterion validity, and growth. The final column of the table notes which stages of research are addressed in each study. We next report studies in examining the effects of teachers' use of mathematics CBMs on student achievement.

Results

Technical Adequacy of Mathematics CBMs for Static Measurement and Progress Monitoring

The 32 identified studies included 25 documents (we refer to them as "studies" though some reports include multiple studies) summarized in Table 1 that reflect an emphasis on research in elementary mathematics (17 studies) relative to either early mathematics (4 studies) or secondary mathematics (4 studies). A similar imbalance is evident in the relative emphasis on Stage 1 studies examining the use of the measures as static indicators (17 studies) in contrast to those addressing Stage 2 (technical characteristics of slopes; 2 studies) or a combination of Stage 1 and Stage 2 (7 studies). While early mathematics studies have relied entirely on measures developed using the robust indicators approach, a mixture of robust indicators and curriculum sampling measures are evident at both the elementary and the secondary levels.

The reliability data include traditional approaches to documenting the measures' internal consistency, test–retest, and alternate form reliability. Some studies (Christ, Johnson-Grohs, & Hintze, 2005; Hintze, Christ, & Keller, 2002) employed generalizability and dependability coefficients. In evaluating the reliability data, we used the traditional benchmark of .80 for typical progress monitoring applications. If the data were to be used for high-stakes decisions, we would suggest applying a more stringent criterion (above .90) for the measures' reliability.

All but one of the studies examined the validity of the measures by investigating concurrent and predictive criterion validity; Shinn and Marston (1985) supported the discriminant validity of the measures by comparing student performance across classification groups. The most common criterion measures are individual and group achievement tests, including (more recently) state standards tests. Readers familiar with the technical adequacy data for elementary CBM reading know that criterion validity coefficients for oral reading fluency are quite high (in the .80 to .90 range). Relations between CBM mathematics measures and criterion variables are generally more modest (in the .50 to .70 range), but quite similar to coefficients for commercially available achievement tests of mathematics (Salvia, Ysseldyke, & Bolt, 2007).

Most studies address student growth by reporting mean weekly slope values representing average increases that might be expected in a student's score each week. To permit comparisons across studies, we converted the growth data in two studies (Chard et al., 20045; Helwig & Tindal, 2002) that were originally reported in other formats into weekly slope rates. One study (Espin et al., 1989) examined cross-sectional means across grade levels on CBM mathematics measures. Rates of growth vary on the measures and appear to be influenced by the complexity of the responses required of students. Research in this area is too immature to offer firm evaluative criteria, but we would assert that measures on which normative growth is more rapid will prove to be more valuable to practitioners than those on which growth is so slow that several weeks' time would be needed to see a single unit of increase in a student's score. Rate of growth, of course, is only one aspect of a measure's sensitivity. The amount of variability around typical growth rates will also affect practitioners' ability to detect changes in student growth. Thus, rate of growth and variability must be examined when selecting progress measures. Research and development efforts on progress measurement must explore both scale properties and methods to control sources of variability that reduce the sensitivity of measures.

Progress Monitoring Measures for Early Mathematics. Research on early mathematics measures has focused entirely on robust indicators, emphasized measures of numeracy, and been conducted exclusively with general education students. Two groups of researchers account for all of the published studies in early mathematics. Clarke and his colleagues (Chard et al., 2005; Clarke & Shinn, 2004) have examined measures such as quantity discrimination and identifying the missing number in a counting sequence with students in kindergarten and first grade. In the quantity discrimination task, students are presented with two numbers and asked to name the larger number. VanDerHeyden and her colleagues

(text continues p. 131

TABLE 1. Technical Adequacy Studies of Mathematics Progress Monitoring Measures

Study	*N*	Grade/ Age	Level[a]	Measure	Type[b]	Reliability	Criterion validity[c]	Growth[d]	Stage(s)[e]
				Early Mathematics					
Chard, Clarke, Baker, Otterstedt, Braun, & Katz (2005)	436 483	K 1	GE GE	• Number identification (NI) • Quantity discrimination (QD) • Missing number (MN)	RI RI RI	—	C: NKT Fall K/1: NI = .65/.56 QD = .55/.45 MN = .69/.61 Spring K/1: NI = .58/.58 QD = .50/.53 MN = .60/.61	K: NI = 1.3 QD = .28 MN = .33 1: NI = .88 QD = .42 MN = .35	1, 2
Clarke & Shinn (2004)	52	1	GE	• Oral counting (OC) • Number identification (NI) • Quantity discrimination (QD) • Missing number (MN)	RI RI RI RI	Alternate form: OC = .99 NI = .89, .93 QD = .93, .92 MN = .83, .78 Test–retest: OC = .93 NI = .97 QD = .96 MN = .84	C: NKT OC = .70 NI = .70 QD = .80 MN = .74 C/P: WJ-AP OC = .60–.64/.72 NI = .63–.65/.72 QD = .71–.79/.79 MN = .68–.69/.72 C/P: Add/Subtr CBM OC = .49–.50/.56 NI = .60–.66/.60–.68 QD = .71–.75/.70–.78 MN = .74–.75/.67–.78	OC = .55 NI = .47 QD = .36 MN = .23	1, 2
VanDerHeyden, Broussard, Fabre, Stanley, LeGendre, & Creppell (2004)	53	Pre-K	GE	• Choose number (CN) • Number naming (NN) • Count objects (CO) • Free count (FC) • Discrimination (D) • Choose shape (CS)	RI RI RI RI RI RI	Alternate form: CN = .83 NN = .87 CO = .87 FC = .71 D = .88 CS = .40	C: Brig/TEMA-2 CN = .57/.52 NN = .47/.39 CO = .44/.49 FC = .56/.19 D = .55/.50 CS = .06/.38		1
VanDerHeyden, Witt, Naquin, & Noell (2001)	107	K	GE	• Circle number (CN) • Write number (WN) • Draw circles (DC)	RI RI RI	Alternate form: CN = .84 WN = .81 DC = .70	C: CIBS-R CN = .61 WN = .44 DC = .52		1

(table continues)

(Table 1 continued)

Study	N	Grade/ Age	Level[a]	Measure	Type[b]	Reliability	Criterion validity[c]	Growth[d]	Stage(s)e
					Elementary Mathematics				
Christ, Johnson-Grohs, & Hintze (2005)	52 52	4 5	GE	Grade-level computation	CS	Generalizability: 64–68% of variability due to measurement error Dependability: Norm-referenced interpretation Low stakes: 1 min High stakes: 4 min Criterion-referenced interpretation Low stakes: 4 min High stakes: 13 min	—		1
Epstein, Polloway, & Patton (1989)	38	9–12 years	SE	• Addition facts (AF) • Subtraction facts (SF)	RI	Test–retest (2 weeks): AF = .85 SF = .80	—		1
Espin, Deno, Maruyama, & Cohen (1989)	159	3–5	GE	Basic facts	RI	Alternate form: Grade 3 = .73 Grade 4 = .93 Grade 5 = .85	C: WRAT Math Comp Grade 3 = .35–.38 Grade 4 = .56 Grade 5 = .48–.50		1, 2
	2,6 0 4	1–6	GE	Basic facts	RI			Slope of means positively accelerating across the grades	
Fuchs, Fuchs, Hamlett, Thompson, Roberts, Kupek, & Stecker (1994)[f]	46 49 45	2 3 4	GE ∑	• MBSP Computation (Comp) • MBSP Concepts & Applications (C & A)	CS CS	Internal consistency: Grade 2 = .98 Grade 3 = .94 Grade 4 = .97	C: CTBS–Math C & A: Grade 2 = .81 Grade 3 = .74 Grade 4 = .79	Weekly slope values (& SD): Comp: Grade 2 = .25 (.25) Grade 3 = .63 (.36) Grade 4 = .70 (.26) C & A: Grade 2 = .40 (.19) Grade 3 = .58 (.30) Grade 4 = .39 (.29)	1, 2
Fuchs, Fuchs, Hamlett, Walz, & Germann (1993) Study 1	18 43 24 21 45 24	1 2 3 4 5 6	GE	MBSP Computation	CS	—	—	Weekly slope (& SD): Grade 1 = .53 (.10) Grade 2 = .20 (.11) Grade 3 = .42 (.26) Grade 4 = .77 (.31) Grade 5 = .70 (.31) Grade 6 = .48 (.38)	2

(table continues)

(Table 1 continued)

Study	*N*	Grade/ Age	Level[a]	Measure	Type[b]	Reliability	Criterion validity[c]	Growth[d]	Stage(s)[e]
Study 2	193 188 195 194 227 211	1 2 3 4 5 6	GE	MBSP Computation	CS	—		Weekly slope (& SD): Grade 1 =.34 (.19) Grade 2 = .28 (.20) Grade 3 = .30 (.23) Grade 4 = .69 (.46) Grade 5 = .74 (.44) Grade 6 = .42 (.49)	
Fuchs, Fuchs, Karns, Hamlett, Dutka, & Katzaroff (2000)	267	2–4	GE	Problem solving performance assessment	CS	Alternate form/Test–retest: .56–.66	C: CTBS—Math and MOAT: .48–.68 C: ITBS PA: .60–.67		1
Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, Hosp, & Jancek (2003)	412	3	GE	Problem solving performance assessment	CS	Internal consistency: .94 Interscorer: .96	C: Ter Nov = .67		1
Fuchs, Hamlett, & Fuchs (1998) Study 1	7 15 20 28	1 2 3 4 5 6	SE	MBSP Computation	CS	Alternate form Grade 2 = .91 Grade 3 = .73 Grade 4 = .92 Grade 5 = .75 Grade 6 = .85			1
Study 2	4 13 12 12 4	1 2 3 4 5	GE	MBSP Computation	CS	Alternate form Grade 4 = .89 Grade 5 = .83 Grade 6 = .93			
Study 3	3 10 19 24 12	6 2 3 4 5	SE	MBSP Computation	CS		C: MCT Grade 2 = .84 Grade 3 = .87 Grade 4 = .84 Grade 5 = .77 C: SAT Math Comp Grade 2 = .93 Grade 3 = .55 Grade 4 = .60 Grade 5 = .59		
Fuchs, Hamlett, & Fuchs (1999)[f]	[46 [49] [45] 51 44	[2] [3] [4] 5 6	GE	• MBSP Computation (Comp) • MBSP Concepts & Applications (C & A)	CS CS	Internal consistency: Grade 5 = .97 Grade 6 = .97	C: CTBS–Total Math C & A: Grade 5 = .80 Grade 6 = .71	Slope values Comp: Grade 5 = .38 Grade 6 = .26	1, 2

(table continues)

(Table 1 continued)

Study	*N*	Grade/ Age	Level[a]	Measure	Type[b]	Reliability	Criterion validity[c]	Growth[d]	Stage(s)[e]
								C & A: Grade 5 = .19 Grade 6 = .12	
Hintze, Christ, & Keller (2002)	67	1–5	GE	• Grade Level single skill computation • Multiple skill computation (cumulative across grades)	RI CS	Generalizability: Single skill: 1 form Multiple skill: 3 forms	—		1
Jitendra, Sczesniak, & Deatline-Buchman (2005)	77	3	GE	• Word problem solving (WPS) • MBSP Computation (Comp)	CS CS	Internal consistency: WPS single probes = .60–.75 WPS aggregation of two forms = .76–.83	C: SAT-9 Problem Solving: WPS = .71 Comp = .49 C: SAT-9 Procedures: WPS = .58 Comp = .64 P: Ter Nov Concepts & Applications WPS = .69 Comp = .38 P: Ter Nov Computation WPS = .62 Comp = .59		1
Shapiro, Edwards, & Zigmond (2005)	120	1–6	SE	• MBSP Computation • MBSP Concepts & Applications (C & A)	CS CS	—	—	Mean slope cross grade: Comp = .38 C & A = .38 Within instructional level: Comp: 1 = .32 2 = .28 3 = .43 4 = .72 5 = .00 C & A: 2 = .36 3 = .37 4 = .44 5 = .52	2

(table continues)

(Table 1 continued)

Study	*N*	Grade/ Age	Level[a]	Measure	Type[b]	Reliability	Criterion validity[c]	Growth[d]	Stage(s)[e]
Shinn & Marston (1985)	209	4–6	GE, SE, and Title 1	• Mixed computation • Multiplication facts • Division facts (Grades 5 and 6 only, *N* = 122)	CS RI RI	—	Discriminant: Significant contrasts in mean performance between GE, SE, and Title 1 students for all measures		1
Thurber, Shinn, & Smolkowski (2002)	207	4	GE	• Basic facts (BF) • Computation (Comp) • Maze reading (Read)	RI CS —	Alternate form: BF = .92 Comp = .90–.92 Read = .87–.88	C: SDMT–Comp: BF = .61–.67 Comp = .54–.59 Read = .54–.57 C: SDMT–Apps: BF = .47–.51 Comp = .36–.42 Read = .55–.58 C: CAT–Comp: BF = .62–.66 Comp = .59–.63 Read = .63–.66 C: CAT–Con/Apps: BF = .50–.55 Comp = .44–.51 Read = .60–.63 C: NAEP Math Items: BF = .45–.52 Comp = .38–.44 Read = .60–.63		1
Tindal, Germann, & Deno (1983)	30 30	4 5	GE	Single skill facts Addition (Add) Subtraction (Subt) Multiplication (Mult) Division (Div)	RI	Alternate form: Grade 4 Add = .72 Subt = .70 Mult = .61 Div = .48 Test–retest (1 week): Grade 5 Add = .87 Subt = .89 Mult = .79 Div = .78	—		1
Tindal, Marston, & Deno (1983)	46	4	GE	Single skill facts Addition (Add) Subtraction (Sub) Multiplication (Mult) Division (Div) Mixed operations (Mix)	RI RI	Alternate form: Add = .87 Sub = .89 Mult = .79 Div = .78 Mix = .93	—		1

(table continues)

(Table 1 continued)

Study	N	Grade/ Age	Level[a]	Measure	Type[b]	Reliability	Criterion validity[c]	Growth[d]	Stage(s)[e]
	30	5	GE			Test–retest (1 week): Add = .72 Sub = .70 Mult = .61 Div = .48			
VanDerHeyden, Witt, & Naquin (2003)	273	1–2	GE	• Grade 1: Addition (answers to 12) • Grade 2: Subtraction (answers to 8)	RI RI	Test–retest (across grades) = .95	—		1
						Secondary Mathematics			
Foegen (2000)	105	6	GE	• Basic facts (BF) • Estimation (Est)	RI RI	Alternate form: Single probe: BF = .85–.92 Est = .75–.85 Mean of two probes: BF = .92–.95 Est = .82–.86	C: Teacher ratings & rankings BF = .47–.66 Est = .51–.62 C: ITBS—Math BF = .45 Est = .59 C: Math grade BF = .52 Est = .46 C: Semester GPA BF = .51 Est = .47	Mean slope: BF = .55 Est = .25	1, 2
Foegen & Deno (2001)	100	6–8	GE	• Basic facts (BF) • Basic estimation (BE) • Modified estimation A (MET-A) • Modified estimation B (MET-B)	RI RI RI RI	Internal consistency: BF = .91–.92 BE = .93 MET-A = .78–.81 MET-B = .77–.81 Alternate form: BF = .79–.80 MET-A = .67–.70 MET-B = .77–.82 Test–retest: BF = .80–.84 BE = .80 MET-A = .67–.73 MET-B = .77–.80	C: CAT—Comp: BF = .63 BE = .56 MET-A = .47 MET-B = .55 C: CAT—Con/Apps: BF = .44 BE = .45 MET-A = .29 MET-B = .55 C: Teacher rating BF = .52 BE = .49 MET-A = .39 MET-B = .51		1

(table continues)

(Table 1 continued)

Study	*N*	Grade/ Age	Level[a]	Measure	Type[b]	Reliability	Criterion validity[c]	Growth[d]	Stage(s)[e]
Helwig, Anderson, & Tindal (2002)	90	8	GE	Concepts	CS	—	C: Computer adaptive state test GE = .80		1
	81	8	SE				SE = .61		
Helwig & Tindal (2002)	117	8	GE	Concepts & applications	CS	Alternate form: .81–.88	State math standards test: .81–.87	Approximate weekly slope .04; 2 of 3 increases across 10-week intervals were statistically significant	1, 2

Note. K = kindergarten; Pre-K = prekindergarten; **Add/Subtr CBM** = Addition or Subtraction curriculum-based measurement task; **Brig** = *Brigance* (Brigance, 1985); **CAT–Con/Apps** = *California Achievement Tests,* Math Concepts and Applications subtest (CTB/McGraw-Hill, 1992); **CAT–Comp** = *California Achievement Tests,* Math Computation subtest (CTB/McGraw-Hill, 1992); **CIBS-R** = *Comprehensive Inventory of Basic Skills–Revised* (Brigance, 1999); **CTBS–Math** = *Comprehensive Test of Basic Skills,* Math Total Battery (Macmillan/McGraw-Hill, 1989); **GPA** = grade point average; **ITBS–Math** = *Iowa Tests of Basic Skills,* Math subtest (Hoover, Hieronymus, Frisbie, & Dunbar, 1993); **ITBS PA** = *Iowa Test of Basic Skills* Performance Assessment (Riverside, 1995); **MCT** = *Math Computation Test;* **MOAT** = *Math Operations and Applications Test;* **NAEP** = *National Assessment of Educational Progress;* **NKT** = *Number Knowledge Test,* Mathematics subtest; **SAT Math Comp** = *Stanford Achievement Test,* Math Computation subtest; **SAT-9** = *Stanford Achievement Test–9th Edition* (Harcourt, Brace & Company, 1996); SDMT–Apps = Stanford Diagnostic Mathematics Test, Applications subtest (Beatty, Gardner, Madden, & Karlsen, 1985); SDMT–Comp = Stanford Diagnostic Mathematics Test, Computation subtest (Beatty, Gardner, Madden, & Karlsen, 1985); **TEMA-2** = *Test of Early Math Ability–2* (Ginsburg & Baroody, 1990); **Ter Nov** = *Terra Nova* (CTB/McGraw-Hill, 2001); **WJ-AP** = *Woodcock Johnson*—Applied Problems subtest (Woodcock & Johnson, 1989); **WRAT Math Comp** = *Wide Range Achievement Test,* Math Computation subtest (Jastak & Wilkinson, 1984).

[a]GE = general education students; SE = special education students. [b]RI = measures representing the robust indicators approach; CS = measures representing the curriculum sampling approach. [c]C: = precedes concurrent criterion validity data; P: = precedes predictive criterion validity data. [d]Any data not reported in weekly slope rates were converted for comparison purposes in this table. [e]1and 2 indicate the research stage(s) represented in the study; Stage 1 research examines technical features of the measures for assessing current (static) performance levels; Stage 2 research examines technical features of the slope when the measures are used in continuous progress monitoring.

[f]Fuchs, Fuchs, Hamlett, Thompson, Roberts, Kupek, and Stecker (1994) includes a subset of the full data reported in Fuchs, Hamlett, and Fuchs (1994). Only the data for fifth- and sixth-grade students (not reported in the published article) are included in the listing for Fuchs, Hamlett, and Fuchs (1994).

(VanDerHeydenet al., 2004; VanDerHeyden, Witt, Naquin, & Noell, 2001) have conducted studies with prekindergarten and kindergarten students using measures such as circling numbers, drawing numbers, and drawing circles. Measures common to both groups of researchers include number naming/identification and counting tasks. The reliability of most of the early mathematics measures has been quite strong, and criterion validity coefficients are generally in the low to moderate range (.40 to .60) for students through kindergarten; stronger correlations have been obtained with first-grade students (above .70) using the quantity discrimination and missing number tasks.

The existing research in early mathematics has focused predominantly on Stage 1 concerns. Only two studies examined students' growth over time. Clarke and Shinn (2004) computed mean weekly growth rates on four measures across a 26-week period and found that an oral counting measure produced the highest level of mean growth. Chard et al. (2005) reported mean scores from fall to spring (a 32-week period) on three different measures. The number identification measure produced substantially higher levels of mean growth than did the other two measures. All the studies in early mathematics have been published since 2000, reflecting the early stage of research in this area. The early mathematics measures with the most supporting evidence at this point include number identification, quantity discrimination, and missing number, though much research remains to be conducted. Future research should examine sensitivity of the measures to student growth and whether teachers' use of data from these measures can improve student achievement.

Progress Monitoring Measures for Elementary Mathematics. The 17 studies addressing measures for elementary mathematics represent a variety of measures, research strategies, and student populations. Over three quarters of the studies have been conducted with general education students and address Stage 1 research issues. The majority of the studies have relied on curriculum sampling measures or a mix of curriculum sampling and robust indicators. When researchers have used curriculum sampling measures, the most common options included either the *Monitoring Basic Skills Progress* measures (MBSP; Fuchs, Hamlett, & Fuchs, 1998, 1999) or a sampling of computation skills within a grade span (Christ et al., 2005). Given the pervasiveness of the MBSP measures in the literature, a more thorough description of these measures is appropriate.

Two types of MBSP mathematics measures are available: (a) Computation and (b) Concepts and Applications, each packaged with a software program that supports teachers' use of the data to make instructional decisions. The Computation measure consists of 30 parallel forms at each of Grades 1 through 6. Fuchs et al. created the measures by selecting problem types representing a proportional sampling of the computation skills within the Tennessee state curriculum at each respective grade level. The Concepts and Applications measure, available for students in Grades 2 through 6, were developed using a similar process, but with attention to applied skills and concepts (i.e., reading charts and graphs, understanding numbers, and solving word problems). Administration times range from 2 to 6 min for the Computation probes and from 6 to 8 min for the Concepts and Applications probes. Two studies examined a problem-solving performance assessment that represented core skills within the grade-level instructional curriculum (Fuchs et al., 2000; Fuchs et al., 2003); a single study relied on a sampling of word problems from the instructional curriculum (Jitendra, Sczesniak, & Deatline-Buchman, 2005). Robust indicators at the elementary level have focused exclusively on basic facts (e.g., Epstein, Polloway, & Patton, 1989). The majority of these measures are 1- or 2-min probes that sample either facts for a single operation or a set of mixed facts representing all four operations.

The reliability of the elementary mathematics measures was generally quite strong, with most coefficients well above .80. Exceptions to this pattern included the measures addressing more complex mathematics tasks such as word-problem solving and problem-solving performance assessments, which have been less reliable. Initial studies involving basic facts (Tindal, Germann, & Deno, 1983; Tindal, Marston, & Deno, 1983) reported somewhat lower reliability coefficients (particularly for multiplication and division) than did other studies involving single skill or mixed operation facts. Across all types of measures, alternate form reliability estimates tended to be lower than test–retest reliability coefficients were, suggesting potential variability in student performance across parallel forms. The generalizability study results confirmed this finding; multiple skill probes representing more complex performance required a greater number of forms to produce a stable estimate of student performance than did single skill probes (Hintze et al., 2002).

The MBSP measures and the word-problem solving measures developed by Jitendra et al. (2005) demonstrated the highest levels of criterion validity, with the majority of the coefficients in the .60 to .80 range. Only two studies addressed the criterion validity of basic facts measures (Espin et al., 1989; Thurber, Shinn, & Smolkowski, 2002); correlation coefficients between these measures and other measures of mathematics computation ranged from .30 to .60. In addition to criterion validity, Shinn and Marston (1985) provided evidence for the validity of mixed computation and basic facts measures by demonstrating significant differences in the performance of three groups of students (general education, Title 1, and special education) in the expected directions for all of the measures.

Student growth on the elementary measures has most often been examined by computing mean weekly slope values for students in general education or special education samples. All but one of the studies (Espin et al., 1989) examining growth was conducted using the MBSP measures; Espin et al. examined mean scores on a basic facts task for students in Grades 1 to 6 and found a pattern of positively accelerating means across grade levels. Most of the growth research on the

MBSP measures has relied on general education students to document normative rates of growth; Shapiro, Edwards, and Zigmond (2005) reported mean slopes for students with disabilities. Weekly growth rates on both the MBSP Computation and Concepts and Applications measures ranged from about one quarter of a point to three quarters of a point per week. Shapiro et al. (2005) obtained mean cross-grade estimates of .38 units per week on both MBSP measures. Within-grade mean slopes revealed more variability, ranging from 0 to .72.

The research base for mathematics CBMs at the elementary level is most extensive with regard to the technical adequacy of the measures as static indicators. While basic facts tasks have been investigated in the largest number of studies, only limited data have documented the criterion validity of these measures, and no data on typical weekly slopes are available to serve as a context against which to index student growth on these measures. The recently developed problem-solving measures and problem-solving performance assessments suffer from a similar lack of data; there is a clear need to examine the viability of these measures for documenting student growth. The most well established measures at the elementary level are the MBSP measures (Fuchs, Hamlett, & Fuchs, 1998, 1999). These measures have empirical data supporting their reliability, criterion validity, and sensitivity in detecting student growth.

Progress Monitoring Measures for Secondary Mathematics. Two groups of researchers account for all of the published studies of mathematics progress monitoring measures for secondary students; research in this area has addressed both Stage 1 (technical adequacy of static measures) and Stage 2 (growth). It is important to note that we did not identify a single study addressing mathematics CBMs for high school students. Foegen and her colleagues (Foegen, 2000; Foegen & Deno, 2001) examined a basic facts measure and an estimation measure, both developed using the robust indicators model. The 3-min estimation measure consists of 40 items, with half requiring computation estimation and half requiring estimation of contextual (word) problems. Students select one of three options as the best estimate for the problem; the alternatives differ by a factor of 10 to encourage students to use estimation and number sense, rather than computation skills, as they respond to the task. Helwig and his colleagues (Helwig, Anderson, & Tindal, 2002; Helwig & Tindal, 2002) created an eighth-grade measure that reflected conceptual understanding of mathematics, rather than computation skills. They investigated a large pool of items likely to predict grade-level mathematics achievement on a statewide test and selected a subset of items that together accounted for the greatest variance in student achievement. Unlike many other forms of CBM tasks, their probes were untimed, though they noted that most students completed the 15-item measures in less than 10 min.

Reliability data for the middle school measures is near or above the .80 threshold, with the exception of one of three variations of an estimation task used by Foegen and Deno (2001). Criterion validity data have included district and state achievement tests, as well as teacher ratings and grade-based measures. The validity coefficients reported by Foegen and her colleagues for the middle school robust indicators ranged from .29 to .66, with the majority of the measures in the .40 to .50 range. Helwig and his colleagues'(Helwig, Anderson, & Tindal, 2002; Helwig & Tindal, 2002) measures demonstrated strong criterion validity with a state mathematics standards test, with coefficients for general education students exceeding .80.

Both groups of researchers have examined student growth observed on the measures. Student performance on the concept measures developed by Helwig et al. increased over four administration periods at levels that were statistically significant, but their practical significance remains questionable (mean weekly growth of .04 units/week). The limited change in student performance may be associated with the small number of items (15) included in this measure. The measures investigated by Foegen et al. produced weekly slope values more comparable to those obtained for the elementary measures (.25 points/week for estimation and .55 for basic facts).

The status of research in secondary mathematics progress monitoring parallels that of early mathematics. All of the secondary research has been published since 2000; clearly, work in this area is also in its infancy. As with the early mathematics measures, considerable research is needed to further establish these measures. Of the three types of potential measures that have been investigated to date (facts, estimation, concepts), the concepts measures are severely limited by the negligible change in student performance over time. While these measures were tremendously successful in predicting student performance on criterion tests (Fuchs et al., 1994), it is as yet undetermined whether they can be successfully used as a progress monitoring measure. As discussed, research on these measures must be conducted to determine whether they are sufficiently sensitive to growth over short periods of time. A pressing need in the area of secondary CBM is for measures appropriate for high school students.

Stage 3 Studies: Teachers' Use of CBM Data to Improve Student Achievement

The 32 studies we identified included 7 that examined Stage 3 issues of instructional utility. In identifying these studies, we focused on those that specifically examined factors associated with teachers' use of mathematics CBM and the relation between these practices and student achievement. As a result, we did not include studies in which mathematics CBM was a component of a larger intervention; this approach has been widely used in Fuchs and colleagues' work on *Peer Assisted Learning Strategies* (PALS). Readers interested in a summary of research including PALS studies are referred to Stecker, Fuchs, and Fuchs (2005).

This final group of studies is summarized in Table 2. We have opted to present these studies chronologically to better capture the developmental nature of work in this area.

Fuchs and her colleagues have conducted all seven studies, and six of the seven investigated the MBSP Computation measure with students in special education. Stage 3 studies in mathematics have examined the viability of enhancements to the original CBM system, the effects of using problem-solving performance assessment data to drive classroom instruction, and the importance of individualizing decisions based on CBM data. In the section that follows, we briefly review these studies.

Two studies examined enhancements to the MBSP software used to support teachers' implementation of CBM. The earliest study of teachers' use of mathematics CBM (Fuchs, Fuchs, & Hamlett, 1989a) compared the effects of setting a static goal with the effects of more dynamic goal adjustment based on student data. Teachers in the dynamic goal condition were prompted by the MBSP software to raise a student's goal when appropriate; teachers in the static goal condition were allowed to change students' goals, but not prompted to do so. The achievement of students in the dynamic goal condition was higher than that in either the static goal condition or a control group. Fuchs, Fuchs, Hamlett, and Stecker (1990) examined the addition of a skills analysis component to the MBSP software program: a chart representing individual students' performance data on specific computation skills. Fuchs et al. (1990) determined that teachers who had access to skills analysis effected greater achievement gains in their students.

Three studies investigated other means of supporting teacher implementation. Fuchs, Fuchs, Hamlett, and Stecker (1991) contrasted the addition of a computerized expert system to the CBM with skills analysis. Results indicated that students of teachers who used the expert system outperformed both those in the no-expert system condition and those in a control group. Allinder and BeckBest (1995) compared the effects of university-based consultation with teachers' self-monitoring of their CBM implementation. Although students in both conditions improved their achievement levels from pre- to posttest, no significant differences were observed between the two conditions. Allinder, Bolling, Oats, and Gagnon (2000) compared CBM with and without self-monitoring to a contrast condition with no CBM. The self-monitoring teachers completed written questionnaires each time the software prompted a change in instruction. Student achievement for the CBM with the self-monitoring group exceeded that of the CBM-only group and that of a contrast group in which teachers were not using CBM. However, no differences in slope were obtained for the CBM groups.

Fuchs, Fuchs, Karns, Hamlett, and Katzaroff's (1999) study of teachers' use of problem-solving performance assessment data examined achievement among three groups of students (those above, at, and below grade level in their mathematics performance) in two conditions. Teachers of students in the performance assessment (PA) condition were provided support in administering, scoring, and adapting instruction based on performance assessment data; students in the control (No PA) condition completed the pre- and posttest assessments only. Although overall achievement favored students in the PA condition, analyses of the data by student performance levels revealed that these differences were evident only for students at or above grade level in their mathematics performance. Students in both the PA and the No PA conditions who were performing below grade level had similar levels of posttest achievement.

Stecker and Fuchs (2000) examined the importance of individualizing instructional changes based on student data. In this study, special education teachers worked with pairs of students, implementing typical CBM procedures and adjusting instruction based on only one of the student's data. Results of the study indicated that while all students demonstrated gains from pre- to posttest, the achievement of target students, upon whose CBM data instructional adjustments were made, exceeded that of their partners receiving similar instruction.

The Stage 3 studies conducted to date provide support for the conclusion that teachers can use data from the MBSP Computation measure to improve student achievement. Data from performance assessment measures were found to produce increased achievement for students at or above grade level, but these same benefits were not obtained for students performing below grade level. It is interesting to note that an enhancement found to be superior to a contrast condition in one study (i.e., CBM with skills analysis in Fuchs, Fuchs, Hamlett, & Stecker, 1990) is subsequently found not to differ from a contrast condition when compared to a further enhanced version (i.e., CBM with skills analysis and expert system support; Fuchs, Fuchs, Hamlett, & Stecker, 1991). These counterintuitive patterns in the Stage 3 research have been discussed by Stecker et al. (2005); further research on the dynamics of expert consultation and research training are warranted. Moreover, there is a clear need to expand the breadth of research on teachers' use of mathematics CBM to improve student achievement. To date, no studies have been conducted in early mathematics or middle school, nor have any of the studies examined robust indicators. We do not yet know whether teachers can use data from robust indicators to improve their students' mathematics achievement.

Discussion

Our review of empirical studies of progress monitoring measures in mathematics revealed that the bulk of the existing research has been conducted at the elementary level, with only limited work in early mathematics and middle school mathematics measures. It is significant to note that we were unable to identify any studies addressing high school mathematics. We situated our review by contrasting measures developed using a robust indicators approach with those developed using

TABLE 2. Studies of Teachers' Use of Mathematics Progress Monitoring Measures to Improve Student Achievement

Study	Sample	Progress monitoring measure	Type	Major contrasts/ Treatment length	Primary dependent measure(s)	Findings
Fuchs, Fuchs, & Hamlett (1989a)	30 SE teachers 60 SE students, Grades 2–9, all with reading, spelling, and math IEP goals	MBSP Computation	CS	CBM w/dynamic goal (CBM DG) CBM w/static goal (CBM SG) Contrast	MCT	CBM DG > CBM SG = Contrast Effect Sizes MCT: CBM DG vs. Contrast = .52 CBM DG vs. CBM SG = .28 CBM SG vs. Contrast = .25
Fuchs, Fuchs, Hamlett, & Stecker (1990)	30 SE teachers 91 SE students, Grades 3–9, all with math IEP goals	MBSP Computation	CS	CBM + skills analysis (CBM + SA) CBM Contrast	MCT-R	CBM + SA > CBM = Contrast Effect Sizes CBM + SA vs. Contrast = .67 CBM + SA vs. CBM = .55 CBM vs. Contrast = .26
Fuchs, Fuchs, Hamlett, & Stecker (1991)	33 SE teachers 66 SE students, Grades 2–8, chronic low math achievement	MBSP Computation	CS	CBM w/skills analysis + expert system (CBM ExS) CBM w/skills analysis (CBM NExS) Contrast	MOT-R CBM slope values	MOT-R CBM ExS > CBM NExS = Contrast Effect Sizes MOT-R: CBM ExS vs. Contrast = .94 CBM ExS vs. CBM NExS = .84 Slope: CBM ExS vs. CBM NExS = 1.11
Allinder & BeckBest (1995)	18 SE teachers 35 SE students, Grades 2–8, all with math IEP goals	MBSP Computation	CS	University based-consultant support for CBM (UBC) Self-monitoring of CBM use (SM)	MCT-R	Students in both conditions improved from pre- to posttest UBC = SM
Fuchs, Fuchs, Karns, Hamlett, & Katzaroff (1999)	16 GE teachers 272 GE students, Grades 2–4; coded as above, at, or below grade level	Problem solving performance assessment	CS of core skills	Performance assessment-driven instruction (PA) No performance assessment-driven instruction (No PA)	Analogous problem solving measures Conceptual underpinnings (CU) Computational applications (CA) Problem solving strategies (PS) Communicative value (COM)	CU, CA, PS, COM: Above grade level: PA > No PA At grade level: PA > No PA Below grade level: PA = No PA Effect Sizes for PA vs. No PA: Above/At/Below grade level CU: 1.34/1.15/0.14 CA: 1.20/0.76/0.26 PS: 1.16/1.08/0.55 COM: 1.35/1.09/0.60

(table continues)

(Table 2 continued)

Study	Sample	Progress monitoring measure	Type	Major contrasts/ Treatment length	Primary dependent measure(s)	Findings
Allinder, Bolling, Oats, & Gagnon (2000)	31 SE teachers 54 SE students, all with math computation IEP goals; average grade level = 4	MBSP Computation	CS	CBM +self monitoring (CBM + SM) CBM Contrast	MCT-R CBM slope	MCT-R CBM + SM > CBM = Contrast Slope CBM + SM = CBM Effect sizes MCT-R: CBM + SM vs. Contrast = .46 CBM + SM vs. CBM = .26 CBM vs. Contrast = .20
Stecker & Fuchs (2000)	22 SE teachers 84 SE students, Grades 2–8, all with math IEP goals (42 pairs with 1 CBM target student and 1 partner student)	MBSP Computation	CS	CBM target student (instructional changes based on this student's data) Partner student (received non-matched instructional changes	MOT-R	Both targets and partners increased from pre- to posttest CBM target students outperformed partners at posttest

Note. CBM = curriculum-based measurement; CS = curriculum sampling; GE = general education; IEP =Individualized Education Program; MBSP = *Monitoring Basic Skills Progress;* MCT = *Math Computation Test;* MCT-R = *Math Computation Test–Revised;* MOT-R = *Math Operations Test–Revised;* SE = special education.

curriculum sampling (Fuchs, 2004). Although research in early mathematics has drawn entirely on the robust indicators approach, studies at the elementary and middle school levels have included measures representing both approaches to development. We did not discern clear patterns in the results favoring one approach over the other. Only three of the studies (Hintze et al., 2002; Shinn & Marston, 1985; Thurber et al., 2002) provided a direct comparison of measures drawn from the two approaches. Results of these studies suggest that both types of measures can discriminate between students of varying mathematics ability levels. Thurber et al. (2002) found that a simple basic facts task (robust indicator) was more strongly correlated with both computation and application outcome measures than was a curriculum-sampled computation measure. A notable finding in this study was that the strength of the correlations for a CBM reading measure exceeded those for either CBM mathematics measure on four of the five criterion measures. Hintze et al. (2002) determined that more complex measures developed through curriculum sampling required that students complete multiple forms to reliably gauge mathematics performance levels.

We also examined the nature of the research that has been conducted, using Fuchs's (2004) three stages. The bulk of the research in CBM mathematics has comprised Stage 1 studies investigating the technical adequacy of the measures as static (one point in time) indicators. Stage 2 studies investigating the technical characteristics of student slopes derived from continuous progress monitoring have been conducted for three early mathematics measures: basic facts, the MBSP measures, and three types of middle school measures. The majority of these studies have reported mean weekly slope values, most of which are less than .5 units per week. Only two measures (MBSP Computation and problem-solving performance assessments) have been investigated in Stage 3 studies exploring the effects of teachers' use of CBM data to improve student achievement. Both of these measures have been used successfully to increase student achievement over comparison conditions; however, patterns in the results suggest that teachers' use of CBM likely needs to be paired with some form of support (consultation or self-monitoring) to produce achievement gains that exceed those obtained in typical practice.

Implications for Practice

Having taken stock of the empirical support for mathematics progress monitoring measures, what recommendations can we offer to teachers? At the early mathematics and middle school levels, the research base is not yet sufficient to support strong recommendations for particular measures. The early mathematics measures used by Clarke, Chard, and colleagues, as well as the middle school measures used by Foegen and colleagues offer minor advantages over other measures at these levels. These measures have adequate initial evidence supporting their technical adequacy as static measures, as well as evidence for their use as indicators of progress. At the elementary level, the MBSP measures, basic facts, and two forms of problem-solving measures all have demonstrated evidence supporting their use as static indicators. For progress monitoring, the MBSP measures have the greatest level of empirical support. The MBSP Computation measure is the only measure with evidence at all three stages of research.

In selecting a measure, teachers must also consider the relative advantages and limitations associated with the curriculum sampling and robust indicators approaches to developing measures. As an example, before teachers select the MBSP Computation measure, they should consider the degree to which the curriculum represented by this measure is similar to the instructional curriculum in their districts. Studies in reading CBM have examined the effects of passages drawn from curriculum programs with varying emphases and determined that these differences do not diminish the technical adequacy of the oral reading fluency measure (Fuchs & Deno, 1991). It is not clear whether this result, based on a robust indicator in reading, would also be found for a curriculum sampling measure in mathematics. We encourage practitioners selecting curriculum sampling measures to consider the degree of "match" between the content of the measures and that of their curriculum. This is particularly important in mathematics given the recent release of *Curriculum Focal Points,* by the National Council of Teachers of Mathematics (2006). This document, which articulates a limited number of specific instructional areas for each grade level (Pre-K through Grade 8), may have substantial influence on future curricula and standards-based assessments, which in turn may affect criterion validity of current curriculum sampling measures.

A final consideration for teachers is the issue of expected weekly growth. Existing research in mathematics CBM suggests that students do not "grow" as rapidly when assessed using mathematics measures as they do in reading when the index is oral reading fluency. Teachers who are choosing a measure to monitor student progress must consider the typical rates at which students improve on the various measures and the importance of being able to observe small increments of improvement in student performance. If a measure has a mean weekly slope of .20, five weeks of data may need to be collected before teachers could expect to see a single unit increase in the student's weekly score. For students who experience severe difficulties in mathematics, this limited degree of sensitivity to growth may not meet teachers' needs. Unfortunately, many of the mathematics measures have produced mean weekly slope values that are less sensitive to growth than teachers might desire. Furthermore, the research conducted to date is limited by small sample sizes and high levels of variability in mean slopes across grade levels, particularly for curriculum sampling measures for which the content of the measure changes with each grade level. We encourage teachers to be aware of the existing research data when selecting among the options for progress monitoring measures in mathematics. We also urge researchers to attend to issues of sensitivity to growth as they develop measures.

Contrasting CBM in Mathematics With CBM in Reading

The search for technically and theoretically appropriate measures, largely resolved in CBM of reading, remains active in mathematics. It is both theoretically interesting and practically important that the expanding research on progress monitoring in mathematics has not resulted in converging answers to the "What to measure?" or "How to measure?" questions posed by Deno and Fuchs (1987). Fuchs's (2004) distinction between robust indicators and curriculum sampling may offer a potential explanation for why converging evidence for answering these questions is more difficult in mathematics than it is in reading. Mathematics curricula are created by sequencing the types of mathematics competence students are to learn within and across grade levels. In the traditional parlance of curriculum developers, this is the "scope and sequence" of the mathematics curriculum. Different types of mathematical competence are addressed: geometry, measurement, and probability, for example. Even when "problem solving" is identified as the central focus of mathematics, extensive diversity exists in the types of problems that students learn to solve across the grades.

Consider reading curricula in contrast to mathematics. From the outset, the goal of reading seems clear and relatively simpler than the goal in mathematics. To read, students must learn to translate printed text into their oral language—to "say and understand" the printed word. Once this initial skill is acquired, growth is discernible through increasing levels of fluency in this performance. The range of vocabulary, the complexity of the language structures, and the variety of background knowledge necessary for adequate comprehension increase, yet the nature of the task remains relatively constant. Unlike mathematics curricula, the scope and sequence of reading curricula less address the fundamental act of reading than they identify embellishments on this fundamental act. This difference is what makes it possible to think that a child might "learn to read" in first grade and become quite competent by third grade—that in the upper grades a child is "reading to learn" rather than "learning to read." Comparable ideas do not make sense in mathematics. It does not fit the structure of our language to say a child "learns to math," and it makes no sense to think that students are no longer learning mathematics in higher grades.

Given the differences in the curriculum design of reading and mathematics, we wonder whether the inherent character of the mathematics curriculum militates against finding a robust indicator that can approximate the level of criterion validity that has been attained for the CBM reading progress measures. Perhaps it is the case that learning mathematics, unlike reading, requires the acquisition of increasingly differentiated types of knowledge as students move to higher levels of development. If the difference between developing competence in mathematics and reading is manifest in the design of those curricula, then we might have to develop progress measures consistent with those differences—that is, a robust indicator approach in reading and a curriculum sampling approach in mathematics.

Are there alternative conceptions of mathematics that might make a robust indicator approach possible or that might enable some combination of the two approaches? We have considered several alternatives. For example, we wonder if core competence in mathematics can be identified that, like reading, is fundamental to success in mathematical thinking and problem solving. Suppose, instead of identifying "mathematics" as the focus, we identify "numeracy"—a concept comparable to literacy. Might a robust indicator of "early numeracy" be developed that is comparable to the CBM oral reading measure? Might growth in an indicator of numeracy be useful in making instructional decisions, as has been the case of growth in passage reading? Perhaps core competence exists and can be measured at different levels of development. Early numeracy measures might be appropriate for Grades 1 to 3, and different types of core competence can be identified, and growth in those competencies might be measured in Grades 4 to 6, 7 to 8, and so on. Finally, like the term *reading*, the term *problem solving* seems central to mathematics. Clearly, the types of problems that students learn to solve change as they move through the curriculum. At the same time, it might be possible to identify a corpus of problems at primary, intermediate, and middle school grade levels that could be repeatedly sampled to index growth in problem solving skill at those grades. Indeed, state standards tests might represent cultural expectations and help to define the parameters of such problems.

Implications for Research

Research is desperately needed in several areas of mathematics progress monitoring. A primary need is for work at the grade levels at which research is extremely limited or nonexistent. The complete absence of any research involving high school mathematics represents a significant gap in our knowledge. Given increasing expectations that all students will master challenging curricula, teachers of students with disabilities and students likely to struggle in mathematics have a critical need for tools to support their efforts to provide effective instructional programs. Rather than focusing on grade levels or general mathematics skills and concepts, as has been common in the elementary and middle school research, progress monitoring tools for high school mathematics will need to address particular content domains, such as algebra and geometry. Although initial work exists at the early grades and in middle school, a similar need exists to expand research for these populations.

Extending research efforts beyond Stage 1 (technical adequacy of the measures as static indicators) is another critical need in the field. The pool of measures about which we have some evidence of their use as progress indicators is substantially smaller than is the pool of measures with evidence for their use as static indicators. Furthermore, Fuchs and her colleagues are the only researchers who have conducted studies investigating teachers' use of progress monitoring data to improve student achievement. Only a single mathematics measure (MBSP Computation) has evidence supporting its use to improve student achievement for students with disabilities. We anticipate that teachers can use other measures with positive effects on student achievement, but until such research is conducted, this remains an empirical question. We urge researchers to move beyond Stage 1 investigations of mathematics measures and explore the important questions associated with students' growth on the measures and achievement outcomes.

The empirical and practical consequences associated with the two approaches to developing measures have yet to be investigated. While the robust indicator approach seems more suited for RIPM's efforts to develop a seamless and flexible system of progress monitoring measures, the bulk of the research to date has focused on curriculum sampling measures. Can either or both of the approaches produce a durable measure, useful across a range of grade levels? Is use of more than one measure necessary to capture mathematics proficiency? Might we need one type of measure to predict important future outcomes, but a different measure, more sensitive to small changes in student performance, to monitor student growth?

These issues and questions invite an increase in research on CBM in mathematics. Mathematics proficiency is receiving national attention as a means of increasing educational and employment outcomes for students and enhancing the quality and competitiveness of the U.S. workforce (Cavanagh, 2006). As more emphasis is placed on improving student outcomes in mathematics, there will be a growing demand for progress monitoring tools that can assess current student learning, predict future performance, and support teachers' efforts to design effective instruction. To meet this need, we and our colleagues are pursuing mathematics progress monitoring research through RIPM to address the areas of greatest importance. We are conducting studies of robust indicators across grade levels, investigating teachers' use of early mathematics measures to improve student achievement, and conducting research in middle schools and in high school algebra. Students deserve to make meaningful and maximal progress in mathematics; educators, then, deserve a comprehensive research base supporting the integrity of their efforts to monitor and increase student growth.

AUTHORS' NOTE

The Research Institute on Progress Monitoring (RIPM) at the University of Minnesota is funded by the U.S. Department of Education, Office of Special Education Programs (Award H324H030003), and partially supported the completion of this work.

REFERENCES

Allinder, R. M., & BeckBest, M. (1995). Differential effects of two approaches to supporting teachers' use of curriculum-based measurement. *School Psychology Review, 24,* 285–296.

Allinder, R. M., Bolling, R. M., Oats, R. G., & Gagnon, W. A. (2000). Effects of teacher self-monitoring implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial & Special Education, 21*, 219–226.

Beatty, L. S., Gardner, E. G., Madden, R., & Karlsen, B. (1985). *The Stanford diagnostic mathematics test* (3rd ed.). San Antonio, TX: Psychological Corp.

Brigance, A. (1985). *Brigance preschool screen.* North Billerica, MA: Curriciulum Associates.

Brigance, A. (1999). *Comprehensive inventory of basic skills* (rev. ed.). North Billerica, MA: Curriculum Associates.

Cavanagh, S. (2006, February 15). White House suggests model used in reading to elevate math skills. *Education Week*. Retrieved September 28, 2006, from http://www7.nationalacademies.org/dbasse/EdWeek_Article_PDF.pdf#search="Math Now

Chard, D., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Instruction, 30*(2), 3–14.

Christ, T. J., Johnson-Grohs, K. N., & Hintze, J. M. (2005). An examination of alternate assessment durations when assessing multiple-skill computational fluency: The generalizability and dependability of curriculum-based outcomes within the context of educational decisions. *Psychology in the Schools, 42,* 615–622.

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.

CTB/McGraw-Hill. (1992). *California achievement tests* (5th ed.). Monterey, CA: CTB Macmillan/McGraw-Hill.

CTB/McGraw-Hill. (2001). *TerraNova.* Monterey, CA: Author.

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

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*, 184–192.

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*(8), 1–16.

Epstein, M. H., Polloway, E. A., & Patton, J. R. (1989). Academic achievement probes: Reliability of measures for students with mild mental retardation. *Special Services in the Schools, 5*(1/2), 23–31.

Espin, C. A., Deno, S. L., Maruyama, G., & Cohen, C. (1989, April). *The Basic Academic Skills Sample (BASS): An instrument for the screening and identification of children at risk for failure in regular education classrooms.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Foegen, A. (2000). Technical adequacy of general outcome measures for middle school mathematics. *Diagnostique, 25*, 175–203.

Foegen, A., & Deno, S. L. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *The Journal of Special Education, 35*, 4–16.

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

Fuchs, L. S., & Deno, S. L. (1991). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58,* 232–243.

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

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989a). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55,* 429–438.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989b). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education, 10,* 43–52.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1990). 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., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*, 617–641.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Thompson, A., Roberts, P. H., Kupek, P., et al. (1994). Technical features of a mathematics concepts and applications curriculum-based measurement system. *Diagnostique, 19*(4), 23–49.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*, 27–48.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., Dutka, S., & Katzaroff, M. (2000). The importance of providing information on the structure and scoring of performance assessments. *Applied Measurement in Education, 13*, 1–34.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal, 36*, 609–646.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., et al. (2003). Explicitly teaching for transfer: Effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 95*~~(2)~~, 293–305.

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1999). *Monitoring basic skills progress: Basic math concepts and applications* [computer program manual]. Austin, TX: PRO-ED.

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1998). *Monitoring basic skills progress: Basic math computation* (2nd ed.) [computer program manual]. Austin, TX: PRO-ED.

Ginsburg, H. P., & Baroody, A. J. (1990). *Test of early mathematics ability* (2nd ed.). Austin, TX: PRO-ED.

Harcourt, Brace, & Company. (1996). *Stanford achievement test* (9th ed.). San Antonio, TX: Harcourt Brace Jovanovich.

Helwig, R., Anderson, L., & Tindal, G. (2002). Using a concept-grounded, curriculum-based measure in mathematics to predict statewide test scores for middle school students with LD. *The Journal of Special Education, 36*, 102–112.

Helwig, R., & Tindal, G. (2002). Using general outcome measures in mathematics to measure adequate yearly progress as mandated by Title I. *Assessment for Effective Intervention, 28*(1), 9–18.

Hintze, J. M., Christ, T. J., & Keller, L. A. (2002). The generalizability of CBM survey-level mathematics assessments: Just how many samples do we need? *School Psychology Review, 31*, 514–528.

Hoover, J. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1993). *Iowa tests of basic skills*. Chicago: Riverside.

Jastak, S., & Wilkinson, G. (1984). *Wide range achievement test–Revised*. Wilmington, DE: Jastak Assessment Systems.

Jitendra, A. K., Sczesniak, E., & Deatline-Buchman, A. (2005). An exploratory validation of curriculum-based mathematical word problem-solving tasks as indicators of mathematics proficiency for third graders. *School Psychology Review, 34,* 358–371.

Jones, E. D., & Krause, J. P. (1988). The effectiveness of data-based instruction by student teachers in classrooms for pupils with mild learning handicaps. *Teacher Education and Special Education, 11,* 9–19.

Macmillan/McGraw-Hill. (1989). *Comprehensive test of basic skills* (4th ed.). Monterey, CA: Author.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence.* Reston, VA: Author.

Reys, B. J., Dingman, S., Sutter, A., & Teuscher, D. (2005). *Development of state-level mathematics curriculum documents: Report of a survey.* Columbia, MO: University of Missouri, Center for the Study of Mathematics Curriculum. Retrieved September 14, 2006, at http://www.mathcurriculumcenter.org/resources/ASSMReport.pdf

Riverside. (1995). *Performance assessments for the Iowa Test of Basic Skills*. Chicago: Author.

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2007). *Assessment in special and inclusive education* (10th ed.). Boston: Houghton Mifflin.

Shapiro, E. S., Edwards, L., & Zigmond, N. (2005). Progress monitoring of mathematics among students with learning disabilities. *Assessment for Effective Intervention, 30(2)*, 15–32.

Shinn, M., & Marston, D. (1985). Differentiating mildly handicapped, low-achieving, and regular education students: A curriculum-based approach. *Remedial and Special Education, 6*, 31–38.

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research & Practice, 15*, 128–134.

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.

Thurber, R. S., Shinn, M. R., & Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review, 31*, 498–513.

Tindal, G., Germann, G., & Deno, S. L. (1983). *Descriptive research on the Pine County norms: A compilation of findings* (Research Report No. 132). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Tindal, G., Marston, D., & Deno, S. L. (1983). *The reliability of direct and repeated measurement* (Research Report No. 109). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

VanDerHeyden, A. M., Broussard, C., Fabre, M., Stanley, J., LeGendre, J., & Creppell, R. (2004). Development and validation of curriculum-based measures of math performance for preschool children. *Journal of Early Intervention, 27*, 27–41.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review, 32*, 204–227.

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.

Woodcock, R. M., & Johnson, M. B. (1989). *Woodcock-Johnson psycho-educational battery–Revised*. Allen, TX: DLM Teaching Resources.