TECHNICAL REPORT #4:

MBSP Concepts & Applications: Comparison of Desirable Characteristics for a Grade Level and Cross-Grade Common Measure

Deanna Spanjers, Cynthia L. Jiban, and Stanley L. Deno

RIPM Year 2: 2004 – 2005

Date of Study: October 2004 - May 2005

May 2008

Produced by the Research Institute on Progress Monitoring (RIPM) (Grant # H324H30003) awarded to the Institute on Community Integration (UCEDD) in collaboration with the Department of Educational Psychology, College of Education and Human Development, at the University of Minnesota, by the Office of Special Education Programs. See progressmonitoring.net.

Abstract

Two studies were conducted to assess the degree to which measures from the Monitoring Basic Skills Program –Concepts and Applications (Fuchs, Hamlett, & Fuchs, 1994) demonstrated desirable characteristics of progress monitoring measures. Two types of Concepts and Applications measures—Grade Level and Common Form—were compared along five characteristics: reliability, validity, growth, time, and scoring. The Common Form measure was designed for use across grade levels. A pilot study conducted with students in grades 2 and 5 examined the reliability of the measures. Results indicated that estimates of test-retest reliability were relatively strong despite the number of forms administered; consistently strong reliability results were obtained across both grade levels when the average of two forms was used. In the second study, participants included students in grades 2, 3, and 5. With the exception of the Common Form for grade 2, the measures reflected moderately strong criterion validity, with standardized test scores and teacher rating as the criterion. Issues related to within- and across-grade growth, time, and scoring are discussed.

MBSP Concepts & Applications: Comparison of Desirable Characteristics for a Grade Level and Cross-Grade Common Measure

In curriculum-based measurement of mathematics proficiency, an array of measures have been used and investigated to varying degrees. Many of these measures represent a sampling of students' yearly curricula in computational skills (e.g., Shinn & Marston, 1985; Skiba, Magnusson, Marston, & Erickson, 1986; Fuchs, Hamlett, & Fuchs, 1990; Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Thurber, Shinn, & Smolkowski, 2002; Hintze, Christ, & Keller, 2002; Evans-Hampton, Skinner, Henington, Sims, & McDaniel, 2002). In addition to those sampling computation, a smaller body of measures have been utilized that sample concepts and/or applications of mathematics from grade level curricula.

Helwig and Tindal (2002) and Helwig, Anderson, and Tindal (2002) investigated use of a concepts and applications measure with eighth-grade students. The measures were untimed, and took approximately 10 minutes for students to complete. Alternate form reliability ranged from .81 to .88; correlations between each single form and the criterion statewide math test ranged from .61 to .87, reflecting strong criterion validity in comparison to many other CBM math measures.

At the elementary school level, curriculum-based measurement of concepts and applications has been limited to measures from the Monitoring Basic Skills Program (MBSP; Fuchs, Hamlett, & Fuchs, 1994), a computer application with 30 measures per grade level at grades 2 through 6. The curricula sampled were Tennessee grade level mathematics standards. Information on reliability and validity of CBM scores from this program is available both in the MBSP Concepts and Applications manual and in one

study in a peer-reviewed journal (Fuchs, Fuchs, Hamlett, Thompson, Roberts, Kubek, & Stecker, 1994). While they appear to describe the same study, the journal article describes a sample of students in grades 2 through 4 (140 students) while the manual includes grades 2 through 6 (235 students). Alternate-form reliability is not reported; instead, internal consistency over time was gauged. The mean score from all odd-numbered measures was correlated with the mean score from even-numbered measures, each mean constituting an aggregation of 10 to 15 scores. These correlations, separated by grade level, ranged from .94 to .98. Criterion validity was studied using the same sample, with the Comprehensive Test of Basic Skills (CTBS)-Computation, -Concepts and Applications, and -Total Math scores serving as criteria. Correlations between a mean of students' last three CBM scores and the criteria ranged from .66 to .81. Correlations between the MBSP Concepts and Applications scores and the MBSP Computation scores ranged from .63 to .90. The weekly slope of growth in student performance across time ranged from .12 to .69 points earned per week.

Because the literature reporting reliability and validity for these measures is limited, further investigation of these same issues for the grade-level specific measures is warranted. One question addressed in this report centers on issues of technical adequacy of grade level MBSP Concepts and Applications measures.

An important limitation of measures based on yearly curriculum sampling is their lack of application to gauging cross-year growth. If the measures are designed to be used by students at certain grade levels, then the measure and the metric change as students move from one grade level to the next. An additional question addressed in this report focuses on an alternate use of the MBSP Concepts and Applications materials within a

measurement scheme designed for gauging cross-year growth. Might MBSP probes taken from a single grade level and re-construed as a common, cross-grade measure prove to have durability in terms of reliably and validly assessing growth in mathematics proficiency?

Purpose

The purpose of this set of two studies was to investigate the degree to which measures from the Monitoring Basic Skills Program –Concepts and Applications (Fuchs, Hamlett, & Fuchs, 1994) demonstrated each of several desirable characteristics of progress monitoring measures. These characteristics include reliability, validity, growth within and across years, efficiency of administration time, and ease of scoring.

Two types of measures were compared: grade level measures and a common form measure for use across grades. Both were taken directly from the MBSP measures which sample a yearly curriculum in concepts and applications, with the third-grade level used as Common Form for participants in all grades.

In the studies described, two types of concepts and applications progress monitoring measures—Grade Level and Common Form—were compared along five characteristics: reliability, validity, growth, time, and scoring. Study 1 was a pilot study that addressed the question of how many Concepts and Applications probes are necessary to administer to students in order to obtain a reliable score. The results of Study 1 guided the design of Study 2, which included a larger sample and addressed all five of the characteristics described.

STUDY 1: RELIABILITY PILOT STUDY

Method

Participants

Participants in the present study were students in an urban elementary school in Minnesota. Students from two second-grade classrooms (n = 36) and two fifth-grade classrooms (n = 29) participated in the study. Demographic information is provided in Table 1.

Table 1
Demographic Information for Study Participants and the School as a Whole

	Sample	School
Ethnicity		
Native American	2%	2%
African American	22%	20%
Asian	11%	8%
Hispanic	5%	6%
White	61%	65%
Receiving special education services	5%	11%
Receiving English Language Learner services	8%	9%
Eligible for free or reduced-price lunch	43%	49%
Female	49%	--[a]

[a]School-wide gender data not available.

Measures

Concepts and Applications probes from the *Monitoring Basic Skills Progress* (MBSP) –*Basic Math* program (Fuchs, Hamlett, & Fuchs, 1999) represented the two Grade Level probes as well as the Common Form probe. Alternate forms were drawn randomly from the MBSP black line masters. A single form of each is included in Appendix A.

Grade-Level probes. Grade Level probes for second-grade students included 18 problems. Skills tested were drawn from the following mathematical areas: counting, number concepts, names of numbers, measurement, charts and graphs, money, fractions, applied computation, and word problems. Grade Level probes for fifth-grade students included 23 problems. Skills tested were drawn from the following areas: numeration, money, measurement, geometry, charts and graphs, fractions and factors, decimals, applied computation, and word problems.

Common Form. Common Form probes were third-grade level MBSP measures, which consisted of 24 problems. Skills were drawn from the same mathematical sub-areas covered in the second-grade measure, plus decimals.

Procedures

All probes were group administered during math class by researchers twice a week for two weeks. During the first week, participants completed three forms of the appropriate Grade Level measure on one day, and three forms of the Common Form measure on another day. During the second week, the probes from week 1 were re-administered, with participants completing each measure exactly one week after the first administration. Order of forms was counterbalanced across participants, with each participant completing forms in the same order during week 1 and week 2 administrations.

Directions were abbreviated versions of those printed in the MBSP manual. These are included as Appendix B. Following the protocol for each level of MBSP measure, administration time was 8 minutes for Grade Level probes for second-grade students, 7

minutes for Grade Level probes for fifth-grade students, and 6 minutes for the Common Form probes (third-grade level).

Probes were administered as paper and pencil tasks. Scores were generated by entering student responses into the Monitoring Basic Skills Program—Concepts and Application software.

Results for Study 1

Descriptive statistics for individual probes and for the average and median of three probes for each week are shown in Tables 1 and 2. Information for Grade Level probes is provided in Table 1, and information for Common Form probes is provided in Table 2.

Table 1
Number of Problems Correct for Grade Level Concepts and Applications: Single Forms and Aggregations of Three Scores

	Week 1			Week 2		
	M	*(SD)*	*n*	*M*	*(SD)*	*n*
Grade 2						
Form A	18.85	(7.91)	33	21.25	(8.34)	36
Form B	16.79	(9.68)	33	21.06	(8.94)	36
Form C	16.82	(8.15)	33	20.33	(8.21)	36
Average	17.55	(8.20)	33	20.88	(8.26)	36
Median	17.91	(8.42)	33	21.08	(8.02)	36
Grade 5						
Form A	11.04	(4.75)	26	13.69	(5.69)	29
Form B	10.04	(4.67)	26	11.28	(5.48)	29
Form C	10.73	(4.99)	26	12.66	(6.19)	29
Average	10.60	(4.11)	26	12.54	(5.01)	29
Median	10.35	(4.47)	26	11.86	(4.98)	29

Table 2
Number of Problems Correct for Common Form Concepts and Applications: Single Forms and Aggregations of Three Scores

	Week 1			Week 2		
	M	*(SD)*	*n*	*M*	*(SD)*	*n*
Grade 2						
Form A	12.63	(6.35)	35	15.67	(7.27)	36
Form B	11.46	(7.60)	35	14.78	(9.60)	36
Form C	8.94	(7.17)	35	11.28	(8.49)	36
Average	11.01	(6.35)	35	13.91	(7.78)	36
Median	11.31	(6.76)	35	14.11	(7.76)	36
Grade 5						
Form A	29.79	(8.88)	28	34.83	(8.08)	29
Form B	30.46	(7.53)	28	34.86	(8.63)	29
Form C	26.39	(9.78)	28	31.62	(9.75)	29
Average	28.88	(8.16)	28	33.77	(8.24)	29
Median	28.96	(8.06)	28	34.21	(7.97)	29

Alternate-form reliability coefficients are provided for each grade and measure—Grade Level and Common Form—in Table 3. Both week 1 and week 2 correlation coefficients are presented.

Table 3
Alternate Form Reliability Estimates for Concepts and Applications

		Week 1	Week 2
Grade 2			
Grade Level:	*r*	.89, .89, .84	.91, .93, .91
	n	33	36
Common Form:	*r*	.73, .75, .69	.81, .76, .74
	n	35	36
Grade 5			
Grade Level:	*r*	.53, .63, .62	.70, .45, .74
	n	26	29
Common Form:	*r*	.81, .79, .84	.78, .86, .78
	n	28	29

Note. $p < .05$ for all correlation coefficients

Alternate form reliability estimates for the Common Form were similar across grades 2 and 5, generally in the .70-.80 range. Estimates for the Grade Level form for grade 2 ranged from .84 to .93 (weeks 1 and 2). Estimates for the Grade Level form for grade 5 ranged from .45 to .74 (weeks 1 and 2). At both grade levels, the easier form (Grade Level for grade 2 and Common Form for grade 5) produced higher initial reliability estimates in Week 1. Improvements in alternate form reliability were generally more substantial in the more difficult form (Common Form for grade 2 and Grade Level form for grade 5).

One week test-retest reliability coefficients are presented for both grades and measures in Table 4. Because three forms were administered, three values for test-retest reliability of single forms and the average of two forms are reported. Test-retest reliability estimates for the average score for three forms, as well as the median score for three forms, are included as well.

Table 4
Test-Retest Reliability Estimates for Concepts and Applications

	1 form	Average: 2 forms	Average: 3 forms	Median: 3 forms
Grade 2				
Grade Level: *r*	.89, .86, .90	.91, .93, .93	.94	.95
n	33	33	33	33
Common Form: *r*	.86, .86, .83	.94, .89, .73	.93	.92
n	35	35	35	35
Grade 5				
Grade Level: *r*	.63, .81, .72	.82, .82, .84	.87	.75
n	26	26	26	26
Common Form: *r*	.80, .73, .82	.85, .87, .54	.88	.88
n	28	28	28	28

Note. $p < .05$ for all correlation coefficients

Test-retest reliability coefficients exceeded .80 in grade 2 for a single form of both types of probes. At grade 5, this benchmark for acceptable reliability was achieved when the average of two forms was used in the analyses. Using the average of three forms produced little, if any improvements over the average of two forms. When the median of three forms was used, the reliability of the Grade Level form decreased for grade 5 students. In general, a single form produced acceptable levels of reliability at grade 2, while the average of two forms was necessary to get acceptable levels of reliability at grade 5.

STUDY 2: RELIABILITY AND VALIDITY

Method

Participants

Participants were students in an urban elementary school in Minnesota. Students from two second-grade classrooms (n = 37), two third-grade classrooms (n = 37), and two fifth-grade classrooms (n = 45) participated in the study. Demographic information is provided in Table 5.

Table 5
Demographic Information for Study Participants and the School as a Whole

	Sample	School
Ethnicity		
Native American	3%	2%
African American	55%	54%
Asian	31%	35%
Hispanic	3%	2%
White	8%	8%
Receiving special education services	12%	12%
Receiving English Language Learner services	33%	32%
Eligible for free or reduced-price lunch	88%	86%
Female	51%	

Measures

Concepts and Applications probes from the *Monitoring Basic Skills Progress* (MBSP)–*Basic Math* program (Fuchs, Hamlett, & Fuchs, 1999) constituted both the Grade Level probes and the Common Form probes in the present study. Measure descriptions are identical to Study 1 and are repeated below. The single forms of each measure included in Appendix A were administered for both Study 1 and Study 2.

Grade Level probes. Grade Level probes for second-grade included 18 problems. Skills tested were drawn from the following mathematical areas: Counting, number concepts, names of numbers, measurement, charts and graphs, money, fractions, applied computation, and word problems. Grade Level probes for fifth-grade students included 23 problems. Skills tested were drawn from the following: Numeration, money, measurement, geometry, charts and graphs, fractions and factors, decimals, applied computation, and word problems.

Common Form. Common Form probes were third-grade level MBSP measures, which consisted of 24 problems. Skills were drawn from the same mathematical sub-areas covered in the second-grade measure, plus decimals.

Northwest Achievement Levels Test (NALT). All students in grades 2-7 who were considered capable of testing in the district where the study occurred were administered an achievement-level version of the NALT Math, a multiple-choice achievement test. Problems included computation and number concepts (e.g. place value), geometry, and applications such as time and measurement. The NALT was administered by district personnel to students in grades 2, 3, and 5 in March.

Minnesota Comprehensive Assessment (MCA). All students in grades 3-5 who were considered capable of testing in Minnesota were administered a grade-level version of the MCA Math, a primarily multiple-choice standards-based achievement test. Areas of math measured were shape, space and measurement; number sense and chance and data; problem solving; and procedures and concepts. Test items do not require direct computation of basic math facts in isolation. The test was designed to measure student achievement in the context of state standards in mathematics. The MCA was administered by district personnel to students in grades 3 and 5 in April.

Teacher ratings. Teachers of participating classrooms completed a form asking them to rate their students' general proficiency in mathematics compared to peers in the same class, on a scale from 1 to 7. Directions included a request that they use the full scale. Teacher ratings of students' math proficiency were collected in fall and again in spring. The Teacher Rating Scale for Students' Math Proficiency is included in Appendix C.

Procedure

All probes were group administered during math class by researchers on two days in the fall and two days in the spring. During the first day, participants completed two forms of the appropriate Grade Level measure, and on the second day, they completed two forms of the Common Form measure. Since the Common Form measure is equivalent to the Grade Level Measure for third-grade students, those students completed the probes in one day in the fall and one day in the spring. The order of forms was counterbalanced across participants in both the fall and the spring administrations. Additional math probes were administered during the same class period as part of a different study.

Directions were abbreviated versions of those printed in the MBSP manual. These are included in Appendix B. Following the protocol for each level of MBSP measure, administration time was 8 minutes for Grade Level probes for second-grade students, 7 minutes for Grade Level probes for fifth-grade students, and 6 minutes for the Common Form probes (third-grade level).

Probes were administered as paper and pencil tasks. Scores were generated by entering student responses into the Monitoring Basic Skills Program—Concepts and Application software.

Results

Table 5 shows the descriptive statistics for the number of problems correct for both Grade Level and Common Form probes across fall and spring.

Table 5
Number of Problems Correct for Common Form and Grade Level Concepts and Applications

	Fall		Spring	
	Common Form *M (SD) / n*	Grade Level Form *M (SD) / n*	Common Form *M (SD) / n*	Grade Level Form *M (SD) / n*
Grade 2				
Form A	11.39 (6.31) / 33	16.15 (7.84) / 33	17.30 (7.34) / 27	22.09 (9.57) / 32
Form B	9.58 (7.05) / 33	13.88 (7.64) / 33	18.89 (10.39) / 28	20.06 (8.25) / 32
Average	10.48 (6.45) / 33	15.02 (7.49) / 33	18.16 (7.85) / 28	21.08 (8.01) / 32
Grade 3				
Form A	19.00 (5.61) / 33		28.63 (12.08) / 30	
Form B	20.58 (9.16) / 33		31.60 (14.17) / 30	
Average	19.79 (6.91) / 33		30.12 (12.55) / 30	
Grade 5				
Form A	30.40 (9.83) / 40	12.90 (7.17) / 39	34.90 (8.59) / 40	15.88 (7.92) / 41
Form B	34.68 (11.03) / 40	13.21 (8.04) / 39	37.65 (8.99) / 40	17.71 (9.48) / 41
Average	32.54 (9.89) / 40	13.05 (7.08) / 39	36.28 (8.23) / 40	16.79 (8.34) / 41

Note: For Grade 3, Grade Level probes were equivalent to Common Form probes

On the fall Common Form probes, completed by all grade levels, second-grade students completed an average of approximately 10 problems correctly (average of two forms). Third-grade students completed about 20 problems correctly, and fifth-grade students completed nearly 33 problems correctly on average. The same trend emerged in the spring, with second grade students completing the fewest number of problems correctly (about 18) and fifth-grade students completing the most number of problems correctly (about 36), based on an average of two forms.

For the fall Grade Level forms for grades 2 and 5, second-grade students completed about 15 problems correctly on average (average of two forms), while fifth-grade students completed about 13.

Reliability

Correlation coefficients for alternate forms administered concurrently are provided in Table 6.

Table 6
Alternate Form Reliability Estimates for Common Form and Grade Level Concepts and Applications

	Fall		Spring	
	Common Form	Grade Level	Common Form	Grade Level
Grade 2	.86	.82	.57	.62
n	33	33	27	32
Grade 3	.74		.83	
n	33		30	
Grade 5	.80	.73	.76	.84
n	40	39	40	41

Note: For Grade 3, Grade Level probes were equivalent to Common Form probes; all correlations significant, $p < .01$

Alternate form reliability estimates were generally similar for Common Form and Grade Level measures across grades. Estimates were in the .70-.80 range, with the

exception of the spring Common Form and Grade Level measures for grade 2, which were.57 and .62, respectively.

Validity

Predictive validity with district and state tests. Table 7 shows predictive validity coefficients for an average score from two probes in fall to spring criterion tests of mathematics achievement, NALT and MCA scores. The table includes coefficients for both Grade Level and Common Form probes.

Table 7
Correlations between Average of Two Fall Probe Scores and Spring NALT and MCA Scores

	NALT		MCA	
	Common Form	Grade Level	Common Form	Grade Level
Grade 2	.38*	.65**		
n	30	31		
Grade 3	.65**		.64**	
n	28		27	
Grade 5	.79**	.80**	.67**	.71**
n	38	36	38	36

Note: For Grade 3, Grade Level probes were equivalent to Common Form probes
* Correlation is significant at the .05 level; ** Correlation is significant at the .01 level

For grade 2, the correlation between the average of two fall probe scores and spring NALT scores was stronger for the Grade Level form (r = .65) than the Common Form (r = .38). For grade 3, correlations between fall probe scores (Common/Grade Level Form) and scores on the NALT and the MCA were both in the .60s. For grade 5, similar correlations were obtained for Common and Grade Level measures. Correlations between probe scores (Common and Grade Level measures) and the two criterion measures ranged from .67-.80.

Concurrent validity with district and state tests. Table 8 shows concurrent validity coefficients for an average score (problems correct) from two probes in spring with the NALT and MCA scores. The table includes coefficients for both Grade Level and Common Form probes.

Table 8
Correlations between Average of Two Spring Probe Scores and Spring NALT and MCA Scores

	NALT		MCA	
	Common Form	Grade Level	Common Form	Grade Level
Grade 2	.53**	.71**		
n	25	30		
Grade 3	.82**		.80**	
n	30		29	
Grade 5	.85**	.84**	.76**	.76**
n	40	41	40	41

Note: For Grade 3, Grade Level probes were equivalent to Common Form probes
** Correlation is significant at the .01 level

For grade 2, the correlation between the average of two spring probes and spring NALT scores was stronger for the Grade Level Form (r = .71) than the Common Form (r = .53). The correlations obtained between the spring probe scores (Common/Grade Level Form) and both NALT and MCA scores were similar for grade 3. For grade 5, correlations were similar for the Common Form and the Grade Level probes. Correlations between probe scores (Common and Grade Level measures) and the two criterion measures ranged from .76-.85.

Concurrent validity with teacher ratings. Table 9 displays validity coefficients between the average score (problems correct) from two probes and teacher ratings for both fall and spring. In each grade, two separate classrooms were included in the study,

so two separate validity coefficients are shown. The table includes coefficients for both Grade Level and Common Form probes.

Table 9
Correlations between Average of Two Probe Scores and Teacher Ratings

	Fall		Spring	
	Common Form	Grade Level	Common Form	Grade Level
Grade 2				
Teacher A	.28	.70**	.58	.76**
n	17	17	12	15
Teacher B	.61*	.76**	.60*	.66**
n	16	16	15	17
Grade 3				
Teacher A	.59*		.72**	
n	17		16	
Teacher B	.70**		.77**	
n	16		13	
Grade 5				
Teacher A	.59**	.59*	.80**	.75**
n	19	18	18	19
Teacher B	.80**	.73**	.71**	.67**
n	21	21	20	20

Note: For Grade 3, Grade Level probes were equivalent to Common Form probes
* Correlation is significant at the .05 level; ** Correlation is significant at the .01 level

For grade 2, all correlations were significant (.01 level) between Grade Level probe scores (average of two forms) and teacher ratings and ranged from .66-.76. Only two of the four correlations between the Common Form probe scores and teacher ratings were significant (.05 level). For grade 3, all correlations were significant (.05 or .01 level) for the Common Form/Grade Level measure and teacher ratings, ranging from .59-.77. For grade 5, all correlations were significant (.05 or .01 level) between both Common Form and Grade Level probes and teacher ratings. Similar correlations were obtained for the Common Form and Grade Level probes.

Growth

Within-year growth. Growth within the school year was assessed by comparing the fall and spring mean scores as shown in Table 5. Note that 17 weeks of school occurred between fall and spring administrations. To examine which measure captured the most growth, differences in means (average of two probes) from fall to spring were standardized by subtracting each score from the mean and dividing by the pooled standard deviation so that effect sizes could be calculated. Effect sizes for growth captured by each measure are presented in Table 10.

Table 10
Within-Grade Growth on Common Form and Grade Level Concepts and Applications Across 17 Weeks of School

	Grade 2	Grade 3	Grade 5
	ES / *n*	*ES* / *n*	*ES* / *n*
Common Form	1.06 / 26	1.43 / 26	0.78 / 36
Grade Level	1.28 / 29		0.79 / 35

Note: Effect sizes (ES) are standardized differences in means from fall (average score from two forms) to spring (average score from two forms); for Grade 3, Grade Level probes are equivalent to Common Form probes

Second-grade students evidenced more growth on the Grade Level measure (ES = 1.28) compared to the Common Form (ES = 1.06), based on the average of two probes. Fifth-grade students demonstrated similar amounts of growth on the Common Form and the Grade Level measures. Students in grade 3 evidenced the most amount of growth (Common Form/Grade Level measure), followed by students in grade 2, and then students in grade 5.

Discussion

One of the purposes of this study was to compare technical soundness of Common Form and Grade Level measures. Comparisons between the two measures at

grades 2 and 5, however, are complicated by differing administration times for the two measure types. Grade Level forms were administered for 8 and 7 minutes for grades 2 and 5, respectively; the Common Form was administered for 6 minutes at all grade levels.

Relative Difficulty of Grade Level and Common Form

The relative difficulty of the Common Form and the Grade Level form presumably differed for students in grades 2 and 5. The Common Form was equivalent to the Grade Level form for grade 3, so for second-grade students, the Grade Level form was presumably an easier task than the Common Form. For fifth-grade students, on the other hand, the Grade Level form was presumably a more difficult task than the Common Form. For grade 5, the mean number of problems correct on the Common Form was greater than on the Grade Level form despite a shorter amount of time to complete the task, which suggests the Common Form was an easier task. For Grade 2, the administration time of the Grade Level form is longer than the Common Form, so no firm comparisons can be made. Yet, the Common Form only contains problems from Grade 3 curriculum, so the assumption can be made that the Common Form was more difficult than the Grade Level form for grade 2.

Reliability

Alternate-form reliability appeared fairly equivalent across grades and form. No apparent trends emerged. All reliability coefficients were significant at the $p < .01$ level and ranged from the upper .50's to the upper .80's. .

Validity

A trend was detected in the results that indicated the Common Form and Grade Level form performed differently in some cases depending on the difficulty level of the

task. As noted earlier, the Grade Level form was presumably an easier task than the Common Form for second-grade students and a more difficult task than the Common Form for fifth-grade students. In grade 2, the Grade Level form, the easier task, generally functioned superior to the Common Form with regard to predictive validity (NALT scores as the criterion) and concurrent validity (teacher ratings and NALT scores as the criterion) based on the average of two forms. The same trend was not apparent in grade 5. The Common Form and Grade Level form appeared to function similarly (in terms of predictive and concurrent validity) in grade 5, based on the average of two forms, despite the Common Form being an easier task than the Grade Level form. It should be noted that the more difficult form for fifth-grade students, the Grade Level form, was still at grade level, while the more difficult form for second-grade students, was above grade level, which may account for the findings.

Estimates of predictive and concurrent validity appeared similar across grades for the average of two forms. The Common Form for grade 2 yielded especially low estimates of predictive validity (NALT scores as the criterion) and concurrent validity (teacher ratings as the criterion), but differences in predictive and concurrent validity estimates may be too small to interpret between the remaining measures—the Grade Level form for grade 2, the Grade Level/Common Form for grade 3, and both forms for grade 5.

Growth

Across-grade growth. The Common Form can be used to gauge growth across grades, because students in grades 2, 3, and 5 completed the Common Form probes. As expected, second-grade students completed the fewest number of problems correctly on

the Common Form (based on the average of two forms), followed by third-grade students, and then fifth-grade students. This trend was apparent in both the fall and spring.

Within-grade growth. Standardized measures of growth suggest similar within-grade growth for the Common Form and Grade Level form for grade 5 based on the average of two forms. For grade 2, the Grade Level form reflected slightly more growth than the Common Form. The Grade Level/Common Form for grade 3 reflected slightly more growth than the measures for grade 2, but the measures for grade 5 reflected the least amount of growth.

Administration Time

Administration times for the measures ranged from 6 to 8 minutes. Although the administration time for the Grade Level form at second grade is 2 minutes longer than the Common Form, the former yielded stronger validity coefficients. In grade 5, the Common Form and Grade Level form appeared to function similarly. Although the Grade Level form is one minute longer than the Common Form, the 1-minute difference in administration time may not be considered significant.

Scoring

Concepts and Applications forms are scored by a computer but require significant time and effort to enter the information. Further, the computer program required to score the measures is currently available only for Macintosh operating systems. The measures cannot be scored by hand due to the complex system by which partial credit is determined.

Conclusion

At all grades, the average score from two Grade Level forms suggested moderately-strong validity and reflected within-grade growth. At grade 5, the Common Form also reflected these characteristics, based on the average of two forms. Criterion validity coefficients for the Grade Level form at grade 2, the Grade Level/Common Form at grade 3, and both forms at grade 5 ranged from .64-.85, with MCA and NALT scores as the criterion. Correlations between scores from these forms and teacher ratings ranged from .59-.80. Utility of the Grade Level measures for grades 2, 3, and 5, and the Common Form for grade 5, is indicated in this study.

For grade 2, the Common Form (average of two forms) did not produce sufficient criterion validity coefficients. Utility of the Common Form for grade 2 is not indicated in the study.

References

Evans-Hampton, T. N., Skinner, C. H., Henington, C., Sims, S., & McDaniel, C. E. (2002). An investigation of situational bias: Conspicuous and covert timing during curriculum-based measurement of mathematics across African American and Caucasian students. *School Psychology Review, 31(4),* 529-539.

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(3),* 617-641.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Thompson, A., Roberts, P. H., Kupek, P., & Stecker, P. M. (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*(1), 27-48.

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1990). *Monitoring basic skills progress: Basic math computation manual.* Austin, TX: Pro-Ed.

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1994). *Monitoring basic skills progress: Basic math manual, concepts and applications.* Austin, TX: Pro-Ed.

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. *Journal of Special Education, 36*(2), 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*(4), 514-528.

Shinn, M., & Marston, D. (1985). Differentiating mildly handicapped, low-achieving, and regular education students: A curriculum-based approach. *Remedial and Special Education, 6*(2), 31-38.

Skiba, R., Magnusson, D., Marston, D., & Erickson, K. (1986). *The assessment of mathematics performance in special education: Achievement tests, proficiency tests, or formative evaluation?* Minneapolis: Special Services, Minneapolis Public Schools.

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(4), 498-513.