



TECHNICAL REPORT #3:

MBSP Computation: Comparison of Desirable Characteristics for a Grade Level and Cross-Grade Common Measure

Cynthia L. Jiban and Stanley L. Deno

RIPM Year 2: 2004 – 2005

Date of Study: October 2004 - May 2005

May 2008

The College of Education
& Human Development

UNIVERSITY OF MINNESOTA



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

The purpose of this study was to compare the technical characteristics of a common form of a previously developed computation measure for progress monitoring with respect to the existing grade level forms of the same measure. Participants in the studies were elementary students in Grades 1, 2, 3 and 5. Researchers used the existing computation measures in Grades 1, 2, and 3 to develop a Common Form of the measure representing content spanning all three grade levels. Students also completed Grade Level forms corresponding to the grade levels in which they were enrolled. Data were collected twice in the fall and once in the spring. Results revealed acceptable levels of alternate form and test-retest reliability, particularly when scores from three measures were averaged. Both forms were found to have moderate to moderately-strong levels of predictive and concurrent validity. The Common Form allowed growth to be measured on a common metric across grade levels, but was not as sensitive to within-grade growth as the Grade Level measure for students in the primary grades. In general, the easier forms of the measures (e.g., Grade Level for grades 1 and 2, Common Form for Grades 3 and 5) tended to produce higher levels of reliability and criterion validity than did the more difficult versions of the measures.

MBSP Computation: 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 curriculum 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). While a variety of these measures are included in studies reporting reliability and validity, the measures investigated do not remain the same across studies in terms of content, sampling procedure, or administration procedures. One set of researchers (Fuchs, Hamlett, & Fuchs, 1990), however, has developed a fixed set of curriculum-sampled computation measures used in a computer application. The Monitoring Basic Skills Progress (MBSP) Basic Math software offers thirty forms of a computation measure at each grade level for grades one through six.

Literature on the technical adequacy of these MBSP Basic Math Computation measures, however, is limited to a small number of studies published within the context of technical manuals for the software. The reliability and validity of scores from the Computation program are described in the MBSP Basic Math Computation Manual (Fuchs et al., 1990). Two studies of alternate form reliability (reported in the technical manual) examine both single scores and aggregations of two scores. In the first, which sampled 79 students with mild disabilities in grades one through six, single form reliability by grade level ranged from $r = .73$ to $r = .92$, with aggregation improving

reliabilities to the $r = .91$ to $r = .96$ range. In the second study, which sampled 48 students without disabilities in grades one through six, single form reliability, only reported for grades four to six, ranged from $r = .83$ to $r = .93$, with aggregated score reliabilities, reported for all six grades, ranging from $r = .93$ to $r = .99$. It should be noted that sample sizes at a single grade level in these studies were as low as four students in some cases.

Criterion validity was studied separately using a sample of 65 students with mild disabilities; the average age of the students was 12.5. Mean scores from multiple CBM forms were correlated with those on the Math Computation Test (Fuchs, Fuchs, Hamlett, & Stecker, 1991), the Stanford Achievement Test (SAT) Concepts of Number subtest, and the SAT Math Computation subtest. Students' progress in mathematics was monitored using materials matching their instructional grade levels (rather than the grade levels in which they were enrolled). When broken out by the grade level which was assigned for appropriate monitoring, these correlations ranged from $r = .49$ to $r = .93$; when all student scores were treated as a unified group, correlations with criteria ranged from $r = .66$ to $r = .83$. The mean CBM score used was based on multiple probes from the MBSP program; however, the exact number of probes was not reported, so it is impossible to consider the degree to which the estimates may have been variably influenced by the number of scores on which they were based.

In two subsequent studies, weekly growth rates on MBSP Computation measures were examined. Fuchs, Fuchs, Hamlett, Walz, and Germann (1993) gave weekly measures to 177 students and, the following year, monthly measures to 1,208 students, ranging from grade 1 to grade 6 and examined mean slopes by grade level. The measures were scored using both digits correct and problems correct methods. When digits correct

was used as the graphed score, weekly slopes ranged from .20 to .77 digits correct per week in year one and from .28 to .74 digits correct per week in year two. Slopes were typically lower when the scoring method used the number of problems correct per week. Another study, reported as a subsample in Fuchs, Fuchs, Hamlett, Thompson, Roberts, Kubek, and Stecker (1994) and as the full grade one to six sample (235 students) in the MBSP Basic Math Concepts and Applications Manual (Fuchs, Hamlett, & Fuchs, 1994), also described mean slopes of growth across time. Here, weekly slopes on the Computation measure ranged from .25 to .70 digits correct per week (Fuchs, Fuchs, Thompson et al., 1994; Fuchs, Hamlett, & Fuchs, 1994).

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 address in this report centers on issues of technical adequacy of grade level MBSP Computation 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 Computation materials within a measurement scheme designed for gauging cross-year growth. Might items taken from the MBSP measures and re-construed as a common, cross-grade form prove to have durability in terms of technical adequacy for reliably and validly assessing growth in mathematics proficiency?

Purpose

The purpose of the present study was to investigate the relative differences in technical adequacy between Grade Level versions of the Monitoring Basic Skills Progress-Computation measures (Fuchs, Hamlett, & Fuchs, 1990) and a researcher-developed Common Form developed by sampling items across multiple grade levels. The specific aspects of technical adequacy that we investigated included reliability, criterion validity, and growth within and across years. The Grade Level measures sampled the annual instructional curriculum in computation at each grade level, while the Common Form represented instructional objectives across multiple grade levels. If the technical adequacy of the Common Form is found to be acceptable, this measure may be advantageous in its ability to model growth across multiple years, an important objective of the research activities of the Research Institute on Progress Monitoring

Method

Participants

This study was conducted in an urban elementary school in Minnesota. Participants were students from two classrooms in each of Grades 1 (n=36), 2 (n=37), 3 (n= 37) and 5 (n=45). Demographics for this sample are compared to those for the whole school in Table 1.

Table 1

Demographic Information for Study Participants and the School as a Whole

	<u>Sample</u>	<u>Schoolwide</u>
Special education services	10%	12%
English Language Learner services	34%	32%
Free / reduced price lunch eligibility	86%	86%
Native American	2%	2%
African American	54%	54%
Asian	34%	35%
Hispanic	3%	2%
White	9%	8%
Female	54%	-- ^a

^aSchoolwide data not available for gender.

Independent Variables

Two forms of the math progress monitoring measures were examined: Grade Level and Common Form. For Grade Level measures, forms of each Grade Level probe were drawn randomly from the MBSP (Fuchs et al., 1990) blackline masters. For Common Form measures, items were drawn randomly from MBSP probes across Grades 1, 2, and 3 and compiled to create forms. Three forms of each type were included in the study. Common Form A and a sample page of each grade level measure is included in Appendix A.

Grade Level forms. Probes at each grade level included 25 problems. Grade 1 tested skills that included addition and subtraction without regrouping. Grade 2 probes assessed 1st grade skills plus addition and subtraction with regrouping. Grade 3 probes assessed the following skills: addition and subtraction with regrouping, basic multiplication, and basic facts division. Grade 5 probes assessed 3rd grade skills, with the

addition of decimals and addition and subtraction of fractions, including those with unlike denominators.

Common Form. Common form probes were a combination of randomly selected problems from first, second, and third grade-level MBSP measures. Each common form was comprised of 50 problems, randomly selected from the pool containing problems from Grades 1, 2, and 3.

Criterion Variables

Northwest Achievement Levels Test (NALT) in Mathematics. All second through seventh graders who were considered capable of testing in the urban district where this study took place 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.

Minnesota Comprehensive Assessment (MCA) in Mathematics. All third and fifth graders 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. Items did not include direct computation of basic math facts in isolation. The test was designed to measure student achievement of state standards in mathematics.

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. Teachers were asked to use the full scale. The Teacher Rating Scale for Students' Math Proficiency is included in Appendix B.

Procedure

Independent variables. Probes were group administered during math class by researchers twice a week for two weeks in fall (the last week in October and the first week in November) and one week in spring (the fourth week in March). In the fall, one week test-retest reliability was gauged for both types of measures. During the first week, students completed three forms of the appropriate Grade Level measure on one day and three forms of the Common Form on the other day. During the second week, the same probes from the first week were administered; each measure was administered exactly one week after the first administration. Order of forms was counterbalanced across participants, with each participant taking forms in the same order during both weeks. In the spring, the three Grade Level and three Common Form probes were re-administered during a single week.

Directions were abbreviated versions of those printed in the MBSP manual, and are included in Appendix C. Following the MBSP protocol guidelines for grade levels, the administration time for Grade Level probes was 2 minutes for students in Grades 1 and 2, 3 minutes for students in Grade 3, and 5 minutes for students in Grade 5. The administration time for the Common Form probes was 2 minutes for students at all grade levels.

All probes were scored by researchers for number of problems correct and number of digits correct. Digits for work within the problem were not scored; only correctly placed digits in the answer itself were counted. Interscorer agreement was calculated on 5% of the probes for each scoring method. Average rates of agreement for the Grade Level probes ranged from 97.6 to 98.9% for problems correct (mean = 98.4%).

For digits correct scoring, the average rates of agreement for each grade level ranged from 98.4 to 99.3% (mean = 98.8%). The average rates of agreement for the Common Form were 99.2% and 99.1% for problems correct and digits correct, respectively.

Criterion measures. Teacher ratings of students' math proficiency were collected in fall and again in spring. The NALT was administered by district personnel to students in Grades 2, 3, and 5 in March. District personnel administered the MCA to students in Grades three and five in April.

Results

Descriptive Statistics

Means, standard deviations, and sample size for each form administered and for aggregations of three forms of each measure type are included in Tables 2 through 5. Results of scoring by problems correct and by digits correct are reported in separate tables, as are data for Grade Level and Common Forms.

The descriptive data reveal the possibility of variability across alternate forms. Mean scores across forms within the same testing period sometimes differed substantially. This result occurred at all grade levels and for both the Grade Level and Common Form measures, regardless of whether problems correct or digits correct scoring was used. Increases from the fall testing (weeks 1 and 2) to spring occurred at all grade levels on both types of measures.

Because the nature of the problems included on each Grade Level measure, direct comparisons cannot be made for students of different grades using the Grade Level data. The Common Form, however, was identical for students in all four grade levels. Data on

this measure revealed consistent increases from one grade level to another at all time periods, regardless of the type of measure.

Table 2

Grade Level Computation, Problems Correct: Means & Standard Deviations for Single Forms and Aggregations of Three Scores

	<u>Week 1</u> <i>M (SD) / n</i>	<u>Week 2</u> <i>M (SD) / n</i>	<u>Spring</u> <i>M (SD) / n</i>
Grade 1			
Form A	7.75 (5.31) / 28	10.32 (5.56) / 28	16.47 (5.74) / 30
Form B	7.29 (4.70) / 28	9.32 (5.91)	15.47 (6.12)
Form C	9.82 (4.06) / 28	12.36 (5.36)	17.30 (5.11)
Average	8.29 (4.10) / 28	10.67 (5.38)	16.41 (5.19)
Median	8.18 (4.35) / 28	10.57 (5.34)	16.43 (5.75)
Grade 2			
Form A	9.74 (3.35) / 31	9.70 (5.10) / 33	13.21 (6.03) / 29
Form B	9.50 (5.45) / 32	11.06 (5.78)	15.48 (7.58)
Form C	9.60 (5.17) / 30	10.55 (5.73)	12.76 (6.24)
Average	9.54 (4.28) / 32	10.43 (5.26)	13.81 (6.28)
Median	9.80 (4.33) / 32	10.55 (5.29)	13.41 (6.23)
Grade 3			
Form A	9.82 (5.08) / 33	10.88 (5.57) / 33	18.03 (6.95) / 30
Form B	7.03 (4.40)	9.58 (5.60)	17.10 (7.24)
Form C	7.03 (4.78)	8.39 (5.71)	16.40 (8.14)
Average	7.96 (4.44)	9.61 (5.43)	17.18 (7.30)
Median	7.85 (4.75)	9.58 (5.59)	17.30 (7.46)
Grade 5 (n=41)			
Form A	4.18 (3.31) / 40	4.72 (3.78) / 40	7.68 (5.56) / 40
Form B	4.45 (3.72)	5.30 (4.27)	8.93 (6.49) / 41
Form C	5.28 (4.03)	6.40 (4.90)	10.07 (6.76) / 41
Average	4.63 (3.51)	5.48 (4.15)	8.83 (6.15) / 41
Median	4.60 (3.63)	5.38 (4.07)	8.90 (6.40) / 41

Table 3

Grade Level Computation, Digits Correct: Means & Standard Deviations for Single Forms and Aggregations of Three Scores

	<u>Week 1</u> <i>M (SD) / n</i>	<u>Week 2</u> <i>M (SD) / n</i>	<u>Spring</u> <i>M (SD) / n</i>
Grade 1			
Form A	8.32 (6.07) / 28	11.39 (6.66) / 28	19.37 (7.18) / 30
Form B	7.68 (5.09)	9.86 (6.61)	17.70 (7.61)
Form C	9.89 (4.18)	12.71 (5.97)	18.97 (6.28)
Average	8.63 (4.50)	11.32 (6.19)	18.68 (6.46)
Median	8.50 (4.64)	11.21 (6.00)	18.73 (7.16)
Grade 2			
Form A	17.00 (5.83) / 31	17.79 (8.69) / 33	24.55 (10.31) / 29
Form B	15.41 (9.13) / 32	17.88 (9.21)	26.34 (12.43)
Form C	16.97 (8.91) / 30	18.48 (9.21)	23.34 (9.82)
Average	16.29 (7.41) / 32	18.05 (8.59)	24.75 (10.31)
Median	16.77 (7.46) / 32	18.15 (8.45)	24.31 (10.11)
Grade 3			
Form A	18.12 (8.70) / 33	19.39 (9.49) / 33	32.47 (11.03) / 30
Form B	14.97 (8.18)	18.91 (10.03)	32.10 (12.38)
Form C	13.30 (8.05)	16.33 (10.49)	31.47 (15.00)
Average	15.46 (7.80)	18.21 (9.59)	32.01 (12.53)
Median	15.61 (7.96)	18.21 (9.91)	32.03 (12.31)
Grade 5			
Form A	24.05 (11.02) / 41	24.47 (12.54) / 40	33.82 (15.26) / 39
Form B	24.88 (11.12) / 41	26.75 (14.18)	37.32 (17.45) / 40
Form C	28.50 (11.76) / 42	32.08 (15.08)	41.90 (18.88) / 40
Average	25.81 (10.72) / 42	27.77 (13.25)	37.46 (16.95) / 40
Median	25.80 (10.80) / 42	28.20 (13.84)	37.26 (17.35) / 40

Table 4

Common Form Computation, Problems Correct: Means & Standard Deviations for Single Forms and Aggregations of Three Scores

	<u>Week 1</u> <i>M(SD) / n</i>	<u>Week 2</u> <i>M(SD) / n</i>	<u>Spring</u> <i>M(SD) / n</i>
Grade 1			
Form A	1.89 (2.62) / 27	2.77 (3.84) / 26	5.10 (4.27) / 30
Form B	5.37 (3.89)	6.73 (5.17)	11.67 (5.98)
Form C	4.15 (3.07)	5.65 (3.93)	8.67 (4.57)
Average	3.80 (2.83)	5.05 (4.09)	8.48 (4.49)
Median	3.67 (2.91)	4.92 (3.88)	8.27 (4.78)
Grade 2			
Form A	6.73 (3.98) / 33	7.67 (4.22) / 33	10.35 (6.18) / 26
Form B	12.52 (5.25)	13.27 (5.74)	17.19 (8.65)
Form C	11.91 (5.63)	12.06 (6.64)	15.65 (8.26)
Average	10.38 (4.48)	11.00 (4.94)	14.40 (7.14)
Median	10.61 (5.09)	11.64 (5.36)	15.04 (7.24)
Grade 3			
Form A	14.53 (8.43) / 34	16.61 (10.88) / 33	20.10 (11.49) / 31
Form B	21.88 (9.53)	24.36 (11.52)	29.60 (13.60) / 30
Form C	20.91 (9.29)	22.15 (11.39)	28.19 (13.32) / 31
Average	19.12 (8.34)	21.04 (10.62)	25.74 (12.53) / 31
Median	20.00 (8.99)	21.33 (10.83)	26.65 (13.48) / 31
Grade 5			
Form A	25.24 (10.13) / 41	27.23 (11.62) / 40	30.87 (10.78) / 39
Form B	32.98 (9.97) / 41	35.95 (10.98)	39.25 (10.64) / 40
Form C	33.74 (10.44) / 42	35.65 (10.38)	39.40 (10.47) / 40
Average	30.63 (9.59) / 42	32.94 (10.61)	36.60 (9.97) / 40
Median	31.69 (9.86) / 42	34.15 (10.98)	37.69 (11.17) / 40

Table 5

Common Form Computation, Digits Correct: Means & Standard Deviations for Single Forms and Aggregations of Three Scores

	<u>Week 1</u> <i>M (SD) / n</i>	<u>Week 2</u> <i>M (SD) / n</i>	<u>Spring</u> <i>M (SD) / n</i>
Grade 1			
Form A	2.44 (3.70) / 27	3.54 (4.92) / 26	7.20 (5.54) / 30
Form B	6.41 (5.10)	7.85 (6.06)	15.63 (8.14)
Form C	4.67 (3.84)	6.27 (4.29)	10.60 (5.56)
Average	4.51 (3.79)	5.88 (4.81)	11.14 (5.82)
Median	4.30 (3.75)	5.88 (4.74)	10.50 (5.80)
Grade 2			
Form A	10.45 (6.20) / 33	11.27 (6.24) / 33	18.04 (9.17) / 26
Form B	17.18 (7.36)	18.27 (8.77)	27.38 (14.01)
Form C	15.39 (7.88)	16.00 (9.40)	22.88 (11.46)
Average	14.34 (6.56)	15.18 (7.46)	22.77 (10.79)
Median	14.52 (6.73)	15.64 (8.07)	23.12 (10.64)
Grade 3			
Form A	22.71 (12.74) / 34	25.88 (16.76) / 33	33.71 (18.35) / 31
Form B	33.35 (14.92)	37.73 (17.50)	46.83 (19.94) / 30
Form C	29.53 (13.66)	31.33 (17.01)	40.19 (19.35) / 31
Average	28.53 (12.86)	31.65 (16.20)	39.93 (18.76) / 31
Median	29.56 (13.87)	31.67 (16.45)	40.60 (19.61) / 31
Grade 5			
Form A	41.41 (16.17) / 41	44.30 (18.99) / 40	50.79 (17.24) / 39
Form B	52.93 (13.35) / 41	55.93 (14.76)	60.20 (14.16) / 40
Form C	48.88 (15.10) / 42	51.80 (15.96)	56.87 (15.51) / 40
Average	47.67 (14.10) / 42	50.68 (15.99)	56.05 (14.72) / 40
Median	48.19 (14.58) / 42	51.03 (16.28)	56.24 (15.45) / 40

Reliability

Correlations coefficients for alternate forms administered concurrently are presented in Tables 6 (Grade Level) and 7 (Common Form). Correlation coefficients for forms administered one week apart (test-retest) in late fall are reported in Tables 8 (Grade Level) and 9 (Common Form).

The data in Tables 6 and 7 include the three correlation coefficients obtained when each of the three measures was correlated with the remaining measures. Separate columns report the results for Weeks 1 and 2 of the fall testing and for the Spring testing. In Grade 1, substantial increases in alternate form reliability were observed from Week 1 to Week 2. This result may indicate that younger children need additional practice activities prior to collecting data on the measures.

Table 6

Grade Level Computation: Alternate Form Reliability

	<u>Week 1</u>	<u>Week 2</u>	<u>Spring</u>
Grade 1			
Problems correct	.65, .60, .68	.89, .90, .85	.78, .73, .77
Digits correct	.68, .60, .69	.91, .92, .87	.79, .72, .78
<i>n</i>	28	28	30
Grade 2			
Problems correct	.72, .68, .69	.89, .83, .84	.85, .91, .79
Digits correct	.73, .70, .72	.89, .84, .83	.85, .90, .82
<i>n</i>	31, 30, 30	33	29
Grade 3			
Problems correct	.78, .87, .78	.89, .89, .91	.92, .97, .94
Digits correct	.80, .88, .78	.87, .87, .89	.92, .96, .93
<i>n</i>	33	33	30
Grade 5			
Problems correct	.84, .86, .88	.89, .89, .88	.94, .92, .95
Digits correct	.82, .86, .88	.80, .89, .87	.95, .90, .95
<i>n</i>	40	40	40, 40, 41

Note: all significant, $p < .01$.

Table 7

Common Form Computation: Alternate Form Reliability

	<u>Week 1</u>	<u>Week 2</u>	<u>Spring</u>
Grade 1			
Problems correct	.75, .69, .61	.89, .84, .81	.68, .78, .76
Digits correct	.74, .75, .64	.87, .80, .81	.69, .76, .76
<i>n</i>	27	26	30
Grade 2			
Problems correct	.72, .78, .69	.67, .56, .79	.86, .65, .84
Digits correct	.75, .80, .74	.74, .68, .83	.84, .67, .87
<i>n</i>	33	33	26
Grade 3			
Problems correct	.82, .65, .81	.86, .83, .82	.94, .89, .91
Digits correct	.84, .71, .86	.86, .85, .83	.93, .88, .92
<i>n</i>	34	33	30, 31, 30
Grade 5			
Problems correct	.81, .79, .87	.86, .94, .89	.79, .85, .84
Digits correct	.81, .82, .87	.87, .93, .89	.80, .87, .84
<i>n</i>	40, 41, 41	40	39, 39, 40

Note: all significant, $p < .01$.

In Tables 8 and 9, we report data for test-retest reliability. The first column of each table reflects the correlation coefficients when scores from a single measure were correlated with scores from the same measure one week later. The remaining columns in the table reflect the effects of averaging either 2 or 3 forms from each week prior to computing the correlation or, in the case of the final column, computing the median of all three forms before computing the correlation. In general, we obtained increases across all grade levels when two forms were averaged, as compared to using scores from a single form. Moving the average of 2 forms to the average of 3 forms resulted in similar or slight increases in the correlation coefficients. Using the median of 3 forms generally resulted in lower levels of reliability than using the average of either 2 or 3 forms.

Table 8

Grade Level Computation: Test-Retest Reliability

	<u>1 form</u>	<u>Average: 2 forms</u>	<u>Average: 3 forms</u>	<u>Median: 3 forms</u>
Grade 1				
Problems correct	.80, .71, .80	.86, .90, .86	.92	.84
Digits correct	.82, .73, .78	.86, .91, .86	.92	.85
<i>n</i>	27	27	27	27
Grade 2				
Problems correct	.75, .83, .70	.87, .84, .88	.91	.88
Digits correct	.71, .83, .65	.86, .78, .86	.88	.84
<i>n</i>	29, 30, 29	30, 29, 30	30	30
Grade 3				
Problems correct	.86, .84, .86	.89, .89, .89	.91	.88
Digits correct	.82, .86, .81	.87, .85, .88	.88	.85
<i>n</i>	31	31	31	31
Grade 5				
<i>Problems correct</i>	.83, .81, .85	.87, .88, .86	.88	.87
<i>Digits correct</i>	.82, .75, .79	.83, .84, .80	.84	.81
<i>n</i>	38	38	38	38

Note: all significant, $p < .01$.

Table 9

Common Form Computation: Test-Retest Reliability

	<u>1 form</u>	<u>Average: 2 forms</u>	<u>Average: 3 forms</u>	<u>Median: 3 forms</u>
Grade 1				
Problems correct	.86, .58, .60	.76, .79, .69	.78	.67
Digits correct	.80, .56, .50	.72, .72, .62	.71	.68
<i>n</i>	25	25	25	25
Grade 2				
Problems correct	.66, .59, .75	.70, .85, .76	.81	.79
Digits correct	.69, .67, .81	.76, .87, .83	.86	.89
<i>n</i>	31	31	31	31
Grade 3				
Problems correct	.83, .88, .88	.90, .90, .93	.93	.89
Digits correct	.85, .89, .88	.90, .91, .93	.93	.93
<i>n</i>	32	32	32	32
Grade 5				
Problems correct	.91, .83, .87	.93, .93, .90	.94	.86
Digits correct	.92, .83, .88	.94, .94, .88	.94	.89
<i>n</i>	39, 39, 40	40	40	40

Note: all significant, $p < .01$.

Validity

Predictive validity with district and state tests. In Tables 10 and 11 correlations with the NALT and MCA are presented. Table 10 includes coefficients for Grade Level probes, and Table 11 includes those for Common Form. Grade 1 students were not administered either criterion test; Grade 2 students were only given the NALT. Results for both Grade Level and Common Form measures revealed relations that were less strong in Grade 2, but in the moderately-strong range (.60 -.75) for students in Grades 3 and 5.

Table 10

Grade Level Computation: Predictive Validity of Average of Three Fall (Week 1) Probe Scores to Spring NALT and MCAs

	<u>NALT</u>	<u>MCA</u>
Grade 1		
Problems correct	-	-
Digits correct	-	-
<i>n</i>		
Grade 2		
Problems correct	.46*	-
Digits correct	.47*	-
<i>n</i>	29	
Grade 3		
Problems correct	.61**	.70**
Digits correct	.63**	.66**
<i>n</i>	29	28
Grade 5		
Problems correct	.75**	.64**
Digits correct	.73**	.61**
<i>n</i>	37	37

Table 11

Common Form Computation: Predictive Validity of Average of Three Fall (Week 1) Probe Scores to Spring NALT and MCA

	<u>NALT</u>	<u>MCA</u>
Grade 1		
Problems correct	-	-
Digits correct	-	-
<i>n</i>		
Grade 2		
Problems correct	.51**	-
Digits correct	.50**	-
<i>n</i>	30	
Grade 3		
Problems correct	.76**	.72**
Digits correct	.78**	.72**
<i>n</i>	29	28
Grade 5		
Problems correct	.75**	.62**
Digits correct	.75**	.61**
<i>n</i>	39	39

Concurrent validity with district and state tests. Tables 12 and 13 show concurrent validity coefficients for an average score from three probes in spring with the NALT and MCA. Results for Grade Level measures are reported in Table 12, and Common Form results are reported in Table 13. The pattern of results obtained for concurrent validity was similar to the findings for predictive validity. Correlations between student performance on the probes and state and district tests were lower for students in Grade 2 (.50 range for Grade Level forms, non-significant for Common Forms), but greater for students in Grades 3 and 5 for both types of forms (.70 - .80).

Table 12

Grade Level Computation: Concurrent Validity of Average of Three Spring Probe Scores to Spring NALT and MCA

	<u>NALT</u>	<u>MCA</u>
Grade 1		
Problems correct	-	-
Digits correct	-	-
<i>n</i>		
Grade 2		
Problems correct	.50**	-
Digits correct	.52**	-
<i>n</i>	27	
Grade 3		
Problems correct	.71**	.76**
Digits correct	.71**	.73**
<i>n</i>	30	29
Grade 5		
Problems correct	.82**	.73**
Digits correct	.79**	.67**
<i>n</i>	41	41

Table 13

Common Form Computation: Concurrent Validity of Average of Three Spring Probe Scores to Spring NALT and MCA

	<u>NALT</u>	<u>MCA</u>
Grade 1		
Problems correct	-	-
Digits correct	-	-
<i>n</i>		
Grade 2		
Problems correct	.37	-
Digits correct	.39	-
<i>n</i>	24	
Grade 3		
Problems correct	.75**	.76**
Digits correct	.77**	.76**
<i>n</i>	31	30
Grade 5		
Problems correct	.84**	.73**
Digits correct	.84**	.73**
<i>n</i>	40	40

Concurrent validity with teacher ratings. Tables 14 and 15 display validity coefficients reflecting the relation between average scores from three probes and teacher ratings. In each grade, two separate classrooms were included in the study, so two separate validity coefficients are shown. Table 14 presents coefficients for the Grade Level scores, and Table 15 reports results for the Common Form. The Grade Level coefficients were statistically significant with the exception of one first grade classroom in the fall. Coefficients for the Grade Level measures ranged from .56 to .79, with the majority in the .60 to .79 range.

Table 14

Grade Level Computation: Concurrent Validity of Average of Three Scores, Teacher Rating as Criterion

	<u>Teacher A</u>		<u>Teacher B</u>	
	<u>Fall</u>	<u>Spring</u>	<u>Fall</u>	<u>Spring</u>
Grade 1				
Problems correct	.31	.70**	.70**	.75**
Digits correct	.34	.72**	.69**	.73**
<i>n</i>	13	13	13	17
Grade 2				
Problems correct	.72**	.58*	.79**	.68**
Digits correct	.71**	.57*	.82**	.63*
<i>n</i>	18	14	14	15
Grade 3				
Problems correct	.68**	.56*	.60*	.58*
Digits correct	.63**	.55*	.54*	.58*
<i>n</i>	18	16	15	13
Grade 5				
Problems correct	.67**	.79**	.75**	.73**
Digits correct	.62**	.62**	.76**	.78**
<i>n</i>	19	19	21	20

Note: Fall averages are from week one scores.

The criterion validity coefficients with teacher ratings were similarly strong for the Common Form as they were for the Grade Level form (see Table 15). Although a larger number of the coefficients for Grade 1 did not achieve statistical significance, the coefficients for the remaining grade levels ranged from .59 to .84, with the majority at or above .65.

Table 15

Common Form Computation: Concurrent Validity of Average of Three Scores, Teacher Rating as Criterion

	<u>Teacher A</u>		<u>Teacher B</u>	
	<u>Fall</u>	<u>Spring</u>	<u>Fall</u>	<u>Spring</u>
Grade 1				
Problems correct	.34	.34	.53	.64**
Digits correct	.41	.43	.60*	.66**
<i>n</i>	13	13	13	17
Grade 2				
Problems correct	.78**	.54	.70**	.60*
Digits correct	.74**	.59**	.66**	.63*
<i>n</i>	18	12	15	14
Grade 3				
Problems correct	.75**	.74**	.62*	.64*
Digits correct	.72**	.74**	.62*	.66*
<i>n</i>	19	17	15	13
Grade 5				
Problems correct	.65**	.77**	.84**	.62**
Digits correct	.63**	.76**	.85**	.64**
<i>n</i>	20	18	22	20

Note: Fall averages are from week one scores.

Growth

Within-year growth. Growth within the school year can be gauged for each measure by comparing fall and spring mean scores depicted in Tables 2 through 5. Note that nineteen weeks of school occurred between the fall (week 1) and the spring administrations. To examine which measure captured the most growth, mean differences from the average of three probes in fall (week 1) to the average of three in spring were standardized so that effect sizes could be calculated. We standardized the difference scores by subtracting each score from the mean and dividing by the pooled standard deviation, which essentially converted each difference score into standard deviation units, or effect sizes. Effect sizes for growth captured by each measure, including both problems correct and digits correct as scoring schemes, are presented in Table 16.

Table 16

Within Grade Growth on Each Measure Across Nineteen Weeks of School, Digits Correct and Problems Correct

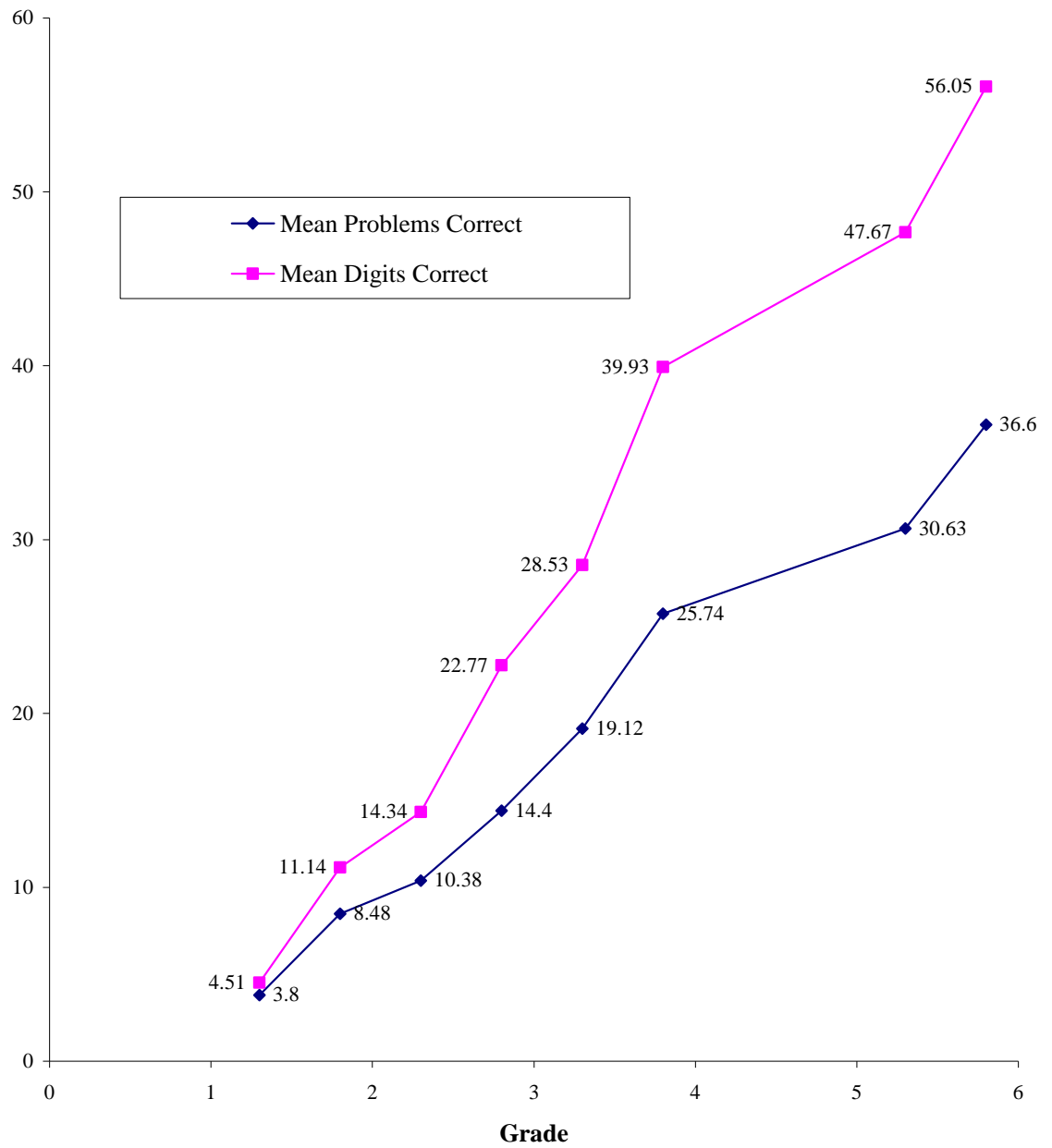
	<u>Grade 1</u> <i>ES</i> / <i>n</i>	<u>Grade 2</u> <i>ES</i> / <i>n</i>	<u>Grade 3</u> <i>ES</i> / <i>n</i>	<u>Grade 5</u> <i>ES</i> / <i>n</i>
Grade Level				
Problems Correct	2.31 / 24	1.13 / 24	2.88 / 28	1.12 / 36
Digits Correct	2.24 / 24	1.27 / 24	2.83 / 28	1.23 / 36
Common Form				
Problems Correct	1.16 / 23	0.69 / 23	1.10 / 28	1.15 / 37
Digits Correct	1.39 / 23	1.04 / 23	1.32 / 28	1.15 / 37

Note: Effect sizes (*ES*) are standardized differences in means from fall (average score from three forms) to spring (average score from three forms).

Across-year growth. Because the content of each of the Grade Level measures is different for each grade, scores from one level cannot be directly compared with another level. Using digits correct as a common metric, we might expect to find across-year growth demonstrated even in the Grade Level measures. Examination of the means across years as shown in Table 2, however, demonstrates that linear growth across years is not gauged by Grade Level measures under either scoring conditions.

In contrast, scores on the Common Form measure can be directly compared across grades. Examining means from either the problems correct (see Table 3) or digits correct (Table 4) metric, growth across years is indicated. This growth in mean scores across grade levels is shown graphically in Figure 1.

Figure 1: Across-Grade Growth on Common Form



Administration Time

Administration time for each measure, as discussed in the Method section above, is presented for direct comparison in Table 17.

Table 17

Administration Times for Grade Level and Common Form Single Probes

	<u>Grade Level</u>	<u>Common Form</u>
Grade 1	2 minutes	2 minutes
Grade 2	2 minutes	2 minutes
Grade 3	3 minutes	2 minutes
Grade 5	5 minutes	2 minutes

Discussion

One of the purposes of this study was to compare technical soundness of a Common Form with the existing Grade Level measures. Throughout, it is important to note that comparisons between Grade Level and Common Forms at Grades 3 and 5 are complicated by the fact of differing administration times for the two measure types. Grade Level forms were administered for three and five minutes, respectively; the Common Form was administered for two minutes at all grades.

Relative Difficulty of Grade Level and Common Form

Relevant to many of the trends observed in this data, we note that the relative difficulty of the two forms differs for each grade level. The Common Form was constructed by sampling items from existing measures for Grades 1 through 3. This implies that for participants in Grades 3 and 5, the Common Form represents an easier task than the Grade Level measure. For students in Grade 1, however, the reverse is true: the Common Form is more difficult than the Grade

Level measure. These inferences are born out by examination of mean performance on the two tasks—even where the Common Form administration time is less than half of that for Grade Level, as at Grade 5, the means on the easier Common Form are higher.

At Grade 2, it is difficult to deduce which form of measure might be more difficult since the Common Form samples from grades both below and above Grade 2 while the Grade Level measure samples curriculum only from Grade 2. The results in this study shed light on this: both had administration times of two minutes, and the mean scores were lower on the Common Form. Common Form, then, represents a more difficult form for participants in Grades 1 and 2, and a less difficult form for Grades 3 and 5.

At Grades 1 and 5, the distribution of scores on the more difficult form was less normal, with larger standard deviations relative to the mean in both cases.

Reliability

Alternate form reliability was, generally, stronger for older participants, regardless of form. While comparisons between Grade Level and Common Form are difficult to make given that there are nine reliability coefficients for each at each grade level, it appears that the two are comparable with Grade Level measures tending to be slightly stronger.

Aggregation of two and three scores generally improves the stability of scores from Week 1 to Week 2. Test-retest reliability coefficients of $r = .80$ or higher for Grade Level measures was achieved using an average score from two forms at Grades 1 and 2 and using just one form score at Grades 3 and 5; $r = .90$ or higher was achieved for Grades 1, 2, and 3 using an average of three scores. On the Common Form, coefficients of $r = .80$ or higher were not achieved for Grade 1, were achieved using an average of three scores for Grade 2, and were

achieved using just one form score at Grades 3 and 5. A coefficient of $r = .90$ or higher was obtained for Common Form at Grades 3 and 5 only, using an average score from two forms.

Overall, test-retest reliability coefficients were greater for the Grade Level measure at Grade 1 but stronger for the Common Form at Grades 3 and 5.

Validity

While the Common Form is at least as good at predicting later performance on the math achievement tests for Grades 2 through 5, the concurrent criterion validity of the Grade Level form is stronger than that of the Common Form at Grade 2. When teacher ratings of math proficiency were the criterion, the Grade Level form showed stronger validity at Grade 1 and possibly Grade 2, while the Common Form was stronger at Grade 3. Differences were too small to interpret for Grade 5.

Growth

Only the Common Form offers a gauge of across-grade growth, by design. When scored as digits correct, the slope of growth across grades is steeper than when the measure was scored using problems correct.

Within-grade growth on measures, after being standardized, was larger on the Grade Level measure for Grades 1, 2, and 3; at Grade 5 the two measures reflected approximately the same amount of growth. The Common Form did not appear to be as sensitive a measure as the Grade Level measures for students in primary grades.

Administration Time

While administration times of all measures were at or below five minutes, a comparison of efficiency at Grade 3 (where the Grade Level form is administered for three minutes) and especially at Grade 5 (where the Grade Level form is administered for five minutes) favors the

Common Form. While a one- or two-minute difference does not seem substantial, these differences are multiplied if several forms must be administered to obtain a reliable score.

Scoring

Reliability of both scoring schemes—problems correct and digits correct—was strong for all measures. While this study did not systematically investigate the time required for each method of scoring, anecdotal evidence from scorers indicates that the problems correct score is more time efficient.

Conclusion

At all grades, the Grade Level forms produced scores which were sufficiently stable for group decision making, which showed moderate to moderately-strong criterion validity, and which reflected significant within-grade growth. At Grade 5, even after aggregating scores on all three forms the grade level measure did not produce scores stable enough for individual decision-making ($r = .90$ or above). However, the reliability of these scores was sufficient to produce criterion validity coefficients ranging from $r = .62$ to $r = .82$ at Grade 5. These results support the utility of the Grade Level measures for all grade levels.

At Grades 1 and 2, the Common Form did not appear to produce sufficiently stable scores for individual decision-making ($r = .90$ or above); furthermore, at these grades it did not capture as much growth as was reflected on the Grade Level measures. The utility of the Common Form at Grades 1 and 2 was not supported by the results obtained in this study.

At Grade 3, the Common Form produced scores with sound reliability and validity, as good or better than those on the Grade Level measure. However, growth reflected on the Grade Level measure was not captured by the Common Form at Grade 3. The utility of the Common Form at Grade 3 was supported for decisions other than gauging within-grade growth.

At Grade 5, the Common Form produced sufficiently sound scores, as or more reliable and valid compared with those from the Grade Level form. Within-grade growth was captured as well by the Common Form as by Grade Level. The utility of the Common Form at Grade 5 was supported in this 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. (1994). *Monitoring basic skills progress: Basic math manual, concepts and applications*. Austin, TX: Pro-Ed.
- Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1990). *Monitoring basic skills progress: Basic math computation manual*. Austin, TX: Pro-Ed.
- 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.

Appendix A:

Grade Level 1 (Sample Page)
Grade Level 2 (Sample Page)
Grade Level 3 (Sample Page)
Grade Level 5 (Sample Page)
Common Form A

Grade Level 1 (Sample Page)

Sheet #1		Computation 1							
Password: ACT									
Name: _____		Date: _____							
A	$\begin{array}{r} 0 \\ + 3 \\ \hline \end{array}$	B	$\begin{array}{r} 7 \\ + 3 \\ \hline \end{array}$	C	$\begin{array}{r} 0 \\ + 7 \\ \hline \end{array}$	D	$\begin{array}{r} 54 \\ + 33 \\ \hline \end{array}$	E	$\begin{array}{r} 7 \\ + 2 \\ \hline \end{array}$
F	$\begin{array}{r} 10 \\ - 0 \\ \hline \end{array}$	G	$\begin{array}{r} 9 \\ + 0 \\ \hline \end{array}$	H	$\begin{array}{r} 0 \\ + 9 \\ \hline \end{array}$	I	$\begin{array}{r} 6 \\ - 0 \\ \hline \end{array}$	J	$\begin{array}{r} 8 \\ - 5 \\ \hline \end{array}$
K	$\begin{array}{r} 10 \\ - 1 \\ \hline \end{array}$	L	$\begin{array}{r} 8 \\ - 1 \\ \hline \end{array}$	M	$\begin{array}{r} 10 \\ - 7 \\ \hline \end{array}$	N	$\begin{array}{r} 1 \\ 7 \\ + 1 \\ \hline \end{array}$	O	$\begin{array}{r} 6 \\ - 2 \\ \hline \end{array}$
P	$\begin{array}{r} 65 \\ + 23 \\ \hline \end{array}$	Q	$\begin{array}{r} 45 \\ - 4 \\ \hline \end{array}$	R	$\begin{array}{r} 5 \\ + 1 \\ \hline \end{array}$	S	$\begin{array}{r} 8 \\ 1 \\ + 0 \\ \hline \end{array}$	T	$\begin{array}{r} 7 \\ - 5 \\ \hline \end{array}$
U	$\begin{array}{r} 8 \\ + 1 \\ \hline \end{array}$	V	$\begin{array}{r} 99 \\ - 8 \\ \hline \end{array}$	W	$\begin{array}{r} 10 \\ - 3 \\ \hline \end{array}$	X	$\begin{array}{r} 9 \\ - 7 \\ \hline \end{array}$	Y	$\begin{array}{r} 9 \\ + 1 \\ \hline \end{array}$

Grade Level 2 (Sample Page)

Sheet #1

Computation 2

Password: AIR

Name: _____ Date: _____

A $\begin{array}{r} 272 \\ - 50 \\ \hline \end{array}$	B $\begin{array}{r} 20 \\ + 24 \\ \hline \end{array}$	C $\begin{array}{r} 6 \\ + 9 \\ \hline \end{array}$	D $\begin{array}{r} 17 \\ - 9 \\ \hline \end{array}$	E $\begin{array}{r} 2 \\ 1 \\ + 8 \\ \hline \end{array}$
F $\begin{array}{r} 16 \\ - 8 \\ \hline \end{array}$	G $\begin{array}{r} 85 \\ - 66 \\ \hline \end{array}$	H $\begin{array}{r} 7 \\ + 9 \\ \hline \end{array}$	I $\begin{array}{r} 13 \\ - 3 \\ \hline \end{array}$	J $\begin{array}{r} 9 \\ + 9 \\ \hline \end{array}$
K $\begin{array}{r} 10 \\ - 6 \\ \hline \end{array}$	L $\begin{array}{r} 70 \\ - 3 \\ \hline \end{array}$	M $\begin{array}{r} 14 \\ + 9 \\ \hline \end{array}$	N $\begin{array}{r} 30 \\ + 6 \\ \hline \end{array}$	O $\begin{array}{r} 8 \\ + 7 \\ \hline \end{array}$
P $\begin{array}{r} 4 \\ + 6 \\ \hline \end{array}$	Q $\begin{array}{r} 31 \\ 23 \\ + 33 \\ \hline \end{array}$	R $\begin{array}{r} 12 \\ - 3 \\ \hline \end{array}$	S $\begin{array}{r} 57 \\ - 8 \\ \hline \end{array}$	T $\begin{array}{r} 3 \\ - 2 \\ \hline \end{array}$
U $\begin{array}{r} 22 \\ - 11 \\ \hline \end{array}$	V $\begin{array}{r} 11 \\ - 6 \\ \hline \end{array}$	W $\begin{array}{r} 4 \\ + 2 \\ \hline \end{array}$	X $\begin{array}{r} 21 \\ + 69 \\ \hline \end{array}$	Y $\begin{array}{r} 14 \\ - 7 \\ \hline \end{array}$

Grade Level 3 (Sample Page)

Sheet #1

Computation 3

Password: ALL

Name: _____ Date: _____

A $\begin{array}{r} 7 \overline{)28} \end{array}$	B $\begin{array}{r} 5 \overline{)35} \end{array}$	C $\begin{array}{r} 98 \\ - 19 \\ \hline \end{array}$	D $1 \overline{)4}$	E $\begin{array}{r} 6 \\ \times 6 \\ \hline \end{array}$
F $\begin{array}{r} 5 \\ \times 4 \\ \hline \end{array}$	G $\begin{array}{r} 400 \\ - 114 \\ \hline \end{array}$	H $\begin{array}{r} 1 \\ \times 0 \\ \hline \end{array}$	I $\begin{array}{r} 73 \\ \times 5 \\ \hline \end{array}$	J $\begin{array}{r} 3 \\ \times 2 \\ \hline \end{array}$
K $\begin{array}{r} 710 \\ + 779 \\ \hline \end{array}$	L $4 \overline{)4}$	M $\begin{array}{r} 97 \\ + 97 \\ \hline \end{array}$	N $\begin{array}{r} 54 \\ \times 4 \\ \hline \end{array}$	O $4 \overline{)32}$
P $\begin{array}{r} 373 \\ - 83 \\ \hline \end{array}$	Q $\begin{array}{r} 8 \\ \times 2 \\ \hline \end{array}$	R $\begin{array}{r} 0 \\ \times 2 \\ \hline \end{array}$	S $\begin{array}{r} 596 \\ + 53 \\ \hline \end{array}$	T $\begin{array}{r} 9 \\ \times 8 \\ \hline \end{array}$
U $\begin{array}{r} 8 \\ \times 3 \\ \hline \end{array}$	V $\begin{array}{r} 1 \\ \times 4 \\ \hline \end{array}$	W $\begin{array}{r} 44 \\ - 37 \\ \hline \end{array}$	X $\begin{array}{r} 308 \\ - 141 \\ \hline \end{array}$	Y $1 \overline{)6}$

Grade Level 5 (Sample Page)

Sheet #1		Computation 5		
Password: ANT				
Name: _____ Date: _____				
A $\frac{1}{3} - \frac{1}{5} =$	B $\frac{2}{3} + \frac{1}{3} =$	C $\begin{array}{r} 53355 \\ - 50568 \\ \hline \end{array}$	D Reduce: $\frac{3}{6} =$	E $6\frac{1}{3} - 2\frac{2}{3} =$
F $\begin{array}{r} 704 \\ \times 69 \\ \hline \end{array}$	G Rename as mixed: $\frac{33}{4} =$	H $6\frac{3}{5} + 3\frac{3}{5} =$	I Rename as improper: $8\frac{5}{7} =$	J $\begin{array}{r} 877 \\ \times 32 \\ \hline \end{array}$
K $\begin{array}{r} 6 \overline{)840} \\ \hline \end{array}$	L $\begin{array}{r} 9 \overline{)634} \\ \hline \end{array}$	M $\frac{1}{7} + \frac{8}{21} =$	N $\begin{array}{r} 58676 \\ 2444 \\ 5231 \\ + 257 \\ \hline \end{array}$	O $\begin{array}{r} 6.22 \\ + 7.7 \\ \hline \end{array}$
P $\begin{array}{r} 17 \overline{)68} \\ \hline \end{array}$	Q $\begin{array}{r} 58320 \\ - 7750 \\ \hline \end{array}$	R $\frac{11}{14} - \frac{5}{7} =$	S $\frac{1}{2} + \frac{1}{5} =$	T $\begin{array}{r} 80 \overline{)83} \\ \hline \end{array}$
U $\begin{array}{r} 2.8 \\ - 1.58 \\ \hline \end{array}$	V Rename as improper: $5\frac{1}{5} =$	W Reduce: $\frac{8}{24} =$	X $\begin{array}{r} 85865 \\ + 81296 \\ \hline \end{array}$	Y Rename as mixed: $\frac{11}{5} =$

Common Form A, page 1

A $\begin{array}{r} 5 \\ \times 2 \\ \hline \end{array}$	B $\begin{array}{r} 7 \\ \times 1 \\ \hline \end{array}$	C $\begin{array}{r} 23 \\ - 18 \\ \hline \end{array}$	D $\begin{array}{r} 6 \\ \times 6 \\ \hline \end{array}$	E $\begin{array}{r} 85 \\ + 16 \\ \hline \end{array}$
F $3 \overline{)24}$	G $\begin{array}{r} 64 \\ + 50 \\ \hline \end{array}$	H $\begin{array}{r} 65 \\ - 57 \\ \hline \end{array}$	I $\begin{array}{r} 9 \\ - 2 \\ \hline \end{array}$	J $\begin{array}{r} 5 \\ - 3 \\ \hline \end{array}$
K $\begin{array}{r} 95 \\ - 3 \\ \hline \end{array}$	L $\begin{array}{r} 912 \\ - 78 \\ \hline \end{array}$	M $\begin{array}{r} 5 \\ - 2 \\ \hline \end{array}$	N $\begin{array}{r} 15 \\ - 9 \\ \hline \end{array}$	O $\begin{array}{r} 978 \\ - 48 \\ \hline \end{array}$
P $\begin{array}{r} 12 \\ - 7 \\ \hline \end{array}$	Q $\begin{array}{r} 3 \\ - 2 \\ \hline \end{array}$	R $\begin{array}{r} 6 \\ - 1 \\ \hline \end{array}$	S $\begin{array}{r} 7 \\ + 7 \\ \hline \end{array}$	T $\begin{array}{r} 4 \\ 6 \\ + 7 \\ \hline \end{array}$
U $\begin{array}{r} 19 \\ \times 8 \\ \hline \end{array}$	V $\begin{array}{r} 13 \\ - 7 \\ \hline \end{array}$	W $\begin{array}{r} 54 \\ \times 4 \\ \hline \end{array}$	X $\begin{array}{r} 7 \\ - 5 \\ \hline \end{array}$	Y $6 \overline{)6}$

Common Form A, page 2

A $\begin{array}{r} 6 \\ + 3 \\ \hline \end{array}$	B $\begin{array}{r} 5 \\ + 5 \\ \hline \end{array}$	C $\begin{array}{r} 3 \\ \times 0 \\ \hline \end{array}$	D $\begin{array}{r} 35 \\ - 21 \\ \hline \end{array}$	E $\begin{array}{r} 60 \\ - 5 \\ \hline \end{array}$
F $\begin{array}{r} 9 \\ + 0 \\ \hline \end{array}$	G $1 \overline{)2}$	H $\begin{array}{r} 9 \\ \times 7 \\ \hline \end{array}$	I $\begin{array}{r} 43 \\ - 27 \\ \hline \end{array}$	J $\begin{array}{r} 7 \\ - 1 \\ \hline \end{array}$
K $\begin{array}{r} 5 \\ - 4 \\ \hline \end{array}$	L $\begin{array}{r} 506 \\ - 324 \\ \hline \end{array}$	M $\begin{array}{r} 98 \\ - 9 \\ \hline \end{array}$	N $\begin{array}{r} 10 \\ - 5 \\ \hline \end{array}$	O $\begin{array}{r} 19 \\ \times 8 \\ \hline \end{array}$
P $\begin{array}{r} 3 \\ + 7 \\ \hline \end{array}$	Q $\begin{array}{r} 11 \\ - 2 \\ \hline \end{array}$	R $\begin{array}{r} 9 \\ - 2 \\ \hline \end{array}$	S $\begin{array}{r} 4 \\ \times 0 \\ \hline \end{array}$	T $\begin{array}{r} 56 \\ + 0 \\ \hline \end{array}$
U $\begin{array}{r} 1 \\ + 3 \\ \hline \end{array}$	V $2 \overline{)8}$	W $\begin{array}{r} 18 \\ - 9 \\ \hline \end{array}$	X $\begin{array}{r} 749 \\ + 76 \\ \hline \end{array}$	Y $\begin{array}{r} 9 \\ + 1 \\ \hline \end{array}$

Appendix C

Abbreviated Directions for MBSP Computation probes**FORM 1**

(Say to the students:)

“Turn to page 1 in your booklet. This is the first orange page. Keep your pencils down. Please listen to these directions and wait until I tell you to begin.

“On these orange pages there are fifty math problems, and you will have ___ minutes to do as many of them as you can.

“Work carefully and do the best you can. When you begin, start at the top left. (*Point.*) Work from left to right. (*Show direction on page.*) When you come to the end of the first page, try the second page. Some problems will be easy for you; others will be harder. When you come to a problem that’s hard for you, skip it, and come back to it later.

“Go through the entire test doing the easy problems. Then go back and try the harder ones. You can get points for getting part of the problem right. So, after you have done all the easy problems, try the harder problems. Do this even if you think you can’t get the whole problem right.

“Remember that you should work across each row, skipping harder problems at first. After you have finished the easier problems on both pages, then go back to the beginning and try the harder ones.

“If you come to a page that says STOP in the middle of it, then STOP. You can flip back and check for problems in the orange section that you have not done. Stay in the orange section only.

“You will have ___ minutes to work. Are there any questions?”

“Ready? Begin.” (*Start stopwatch as you say BEGIN.*)

(After ___ minutes, say:)

“Stop. Thank you, put your pencil down. Turn to page ___ in your booklet.”

FORM 2

(Say to the students:)

“Now you will try another set of math problems, this time on green pages. Work from left to right. Remember, when you come to a problem that’s hard for you, skip it, and come back to it later. After you have done all the easy problems, try the harder problems. Do this even if you think you can’t get the whole problem right.

“If you come to a page that says STOP in the middle of it, then STOP. You can flip back and check for problems in the green section that you have not done. Stay in the green section only.

“Ready? Begin.” (*Start stopwatch as you say BEGIN.*)

(*After ___ minutes, say:*)

“Stop. Thank you, put your pencil down. Turn to page ___ in your booklet.”

FORM 3

(Say to the students:)

“Now you will try one last set of math problems, this time on white pages. Work from left to right. Remember, when you come to a problem that’s hard for you, skip it, and come back to it later. After you have done all the easy problems, try the harder problems. Do this even if you think you can’t get the whole problem right.

“Stay in the white pages only.”

“Ready? Begin.” (*Start stopwatch as you say BEGIN.*)

(*After ___ minutes, say:*)

“Stop. Thank you, put your pencil down. Please close your booklet.”