Research Institute on Progress Monitoring

TECHNICAL REPORT #14:

Establishing Technically Adequate Measures of Progress in Early Mathematics

Erica S. Lembke, Anne Foegen, Tiffany A. Whittaker, David Hampton, and Cynthia L. Jiban

RIPM Year 3: 2005 – 2006

Date of Study: November 2005 – May 2006

August 2008

The College of Education & Human Development

UNIVERSITY OF MINNESOTA

IDEAs that Work

U.S. Office of Special Education Programs

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 examine the use of three early numeracy measures to monitor the mathematics progress of students across time. One hundred and seven Kindergarten and Grade 1 students were administered Quantity Discrimination, Number Identification, and Missing Number measures once each month for 7 months. Alternate form reliability was adequate for instructional decision-making, while criterion validity coefficients comparing the early numeracy measures to teacher judgment of student proficiency in mathematics and students' performance on a district-administered standardized test were lower than those observed in previous research. We used Hierarchical Linear Modeling at each grade level to examine the technical adequacy of the three measures for modeling growth across time. All measures produced growth rates that were significant across time, for each grade level, with linear growth observed for the Number Identification measure only.

Establishing Technically Adequate Measures of Progress in Early Mathematics

Current reform efforts in mathematics have focused on establishing rigorous standards across grade levels for students with and without disabilities (National Council of Teachers of Mathematics, 2000). Simultaneously, government requirements such as those contained in the Individuals with Disabilities Education Improvement Act of 2004 and the No Child Left Behind Act of 2001, have increased pressure on educators to be accountable for the achievement of all students regardless of disability (Miller & Hudson, 2007). Although students' difficulties in reading have traditionally received the most attention, educators are now focusing on mathematics as they attempt to meet federal and state mandates for improving student achievement. The 2005 National Assessment of Educational Progress (Perie, Grigg, & Dion, 2005) revealed that while the mathematics performance of 35% of fourth-graders was classified as proficient or above, 21% of the nation's fourth graders demonstrated mathematics performance levels considered below basic.

For students in early elementary grades, later proficiency in mathematics is fundamental for several reasons. First, mathematics knowledge supports important life skills such as managing personal and household finances and purchasing and maintaining a home and automobile (Minskoff & Allsopp, 2003). Second, many of today's jobs require knowledge of basic mathematics, conceptual understanding of mathematical processes, and proficient problem-solving strategies (Montague, 1996). Third, mathematics proficiency is integral to success in other subject areas such as physical science, economics, and computer literacy (Minskoff & Allsopp, 2003). Finally, performance on high stakes assessments in mathematics is often used as a benchmark in determining advancement to the next grade level (Minskoff & Allsopp, 2003).

As educators and administrators strive to improve the mathematics performance of their students, they can turn to several sources to inform their decisions. Professional organizations, such as the National Council of Teachers of Mathematics (NCTM, 2000) offer guidance through their principles and standards for mathematics curricula and instructional methods. Reviews of the literature and studies on effective mathematics interventions by researchers such as Swanson (1999), Fuchs and Fuchs (2001), and Kroesbergen and Van Luit (2003) provide recommendations specific to students who are struggling in mathematics. As teachers develop a repertoire of potential interventions, they need data to guide their decisions about when an intervention should be implemented, whether the intervention is effective, and whether a modification needs to be made. One technically adequate method for determining whether instruction is effective and whether students are making progress is Curriculum-Based Measurement (CBM; Deno, 1985).

Research in CBM has sought to establish an efficient measurement system that produces accurate, meaningful data for determining proficiency and can assist teachers in determining the effectiveness of their instruction (Deno, Fuchs, Marston, & Shin, 2001). CBM studies illustrate strong criterion validity with widely used commercial assessments, informal measures, and teacher perceptions of student competence (Deno et. al., 2001). Research has determined that when teachers use student progress monitoring, students learn more, teacher decision making improves, and students become more aware of their own performance (Safer & Fleischman, 2005). Two of the primary uses of CBM are for screening and progress monitoring.

Screening of students is important in order to ascertain achievement levels, identify students who are at-risk for difficulties, make placement decisions, and evaluate program effectiveness. Student progress monitoring allows teachers to map student outcome data on a routine basis (i.e., weekly, biweekly, monthly), measuring student growth over time. While research on the use of CBM for screening and progress monitoring has been plentiful in the area of reading (see Deno, Mirkin, & Chiang, 1982; Good, Gruba, & Kaminski, 2001; Stecker, Fuchs, & Fuchs, 2005), research in mathematics has been less abundant, with only a limited number of studies on the use of CBM for screening and progress monitoring in early numeracy (Foegen, Jiban, & Deno, 2007). Studies of early numeracy measures have focused primarily on their use for screening (Baker, Gersten, Flojo, Katz, Chard, & Clarke, 2002; Clarke & Shinn, 2004; Fuchs, Fuchs, Compton, Hamlett, & Seethaler, 2007). It is equally important to investigate the effects of using early mathematics measures to monitor student progress over time (Fuchs, et al., 2007), determining which measures may be effective for both screening and progress monitoring, as well as those that may be useful for only one purpose (Fuchs & Fuchs, 2001).

In an initial study to examine the potential of early mathematics CBM measures for screening, Clarke and Shinn (2004) studied measures that included number identification (NI), quantity discrimination (QD), missing number (MN), and oral counting (OC). They examined the measures' reliability and validity for early identification and the sensitivity of the measures for use in formative evaluation. Clarke and Shinn's results demonstrated that QD, OC, and NI appeared to be the most reliable measures with coefficients over r = .90 for alternative form reliability. Concurrent validity correlations ranged from r = .74 to .79 when compared to the Woodcock-Johnson Applied Problems subtest (WJ-AP; Woodcock & Johnson, 1989), Number Knowledge Test (NKT; Griffin, Case, & Siegler, 1994), and Mathematics Curriculum-Based Measurement (M-CBM) first grade computation probes. Predictive validity showed similar results with QD displaying the highest coefficients. Chard, Clarke, Baker, Otterstedt, Braun, and Katz (2005) reported on the early mathematics measures' sensitivity to growth by examining gains made from fall to spring**.** The authors commented that only the NI task showed considerable change from fall to spring and that the QD and MN measures showed less change across time.

Fuchs et al. (2007) conducted a study to explore methods for screening first grade students needing attention in mathematics by utilizing a Response to Intervention (RTI) model. These researchers sought to summarize studies examining math screening in Kindergarten and Grade 1, while exploring progress monitoring in mathematics for these students (Fuchs et. al., 2007). Screening measures consisting of fact retrieval, CBM computation, Number Identification/Counting, and CBM Concepts/Applications (Fuchs, et al., 2007) were administered to 667 Grade 1 students, followed by weekly progress monitoring consisting of CBM computation and Number Identification/Counting. Results of this study suggested that multi-skill screening probes should be utilized for screening purposes. Fuchs et al., (2007) also stated that Number Identification/Counting should receive additional research to examine its utility as an adequate screening tool. In the examination of progress monitoring measures, these researchers found similar support for multi-skill probes over the simpler, single skill screening measures. The results from Fuchs and colleagues (2007) suggest that measures that are useful in screening of math proficiency may not be equally suitable for progress monitoring.

Lembke and Foegen (2007) examined the reliability and validity of several early numeracy measures (Number Identification, Quantity Array, Quantity Discrimination, and Missing Number) for screening students in Kindergarten and Grade 1. Over 300 students across two states completed the early mathematics measures at one or two time points during the year. Results of their study indicated the strongest reliability coefficients for both alternate form and test retest reliability were obtained for the Number Identification, Quantity Discrimination, and

Missing Number tasks for both Kindergarten and Grade 1 students, with most correlations in the mid to high .80s. With respect to validity, findings were generally parallel to those of the reliability studies, with moderate to strong concurrent and predictive validity coefficients for the Number Identification, Quantity Discrimination, and Missing Number tasks. The Quantity Array measure consistently produced the lowest correlations across grades.

Extending work examining the use of early numeracy measures for screening, the purpose of this study was to examine the technical adequacy of early numeracy CBM for monitoring student growth or progress across time. At seven points during a school year, we administered early numeracy measures deemed technically adequate in previous research, including Number Identification, Quantity Discrimination, and Missing Number. We examined alternate form reliability, criterion validity, the ability of the measures to model growth across time, and typical growth rates.

Method

Setting and Participants

Participants in this study were 77 Kindergarten and 30 Grade 1 students from the classrooms of 6 teachers in an elementary school in a rural Midwestern town with a total district enrollment of 1,196. Fifty-seven students were male and fifty were female. Ninety-four percent of the students were Caucasian, while 5% were African-American, and 1% were Hispanic. Twenty-five percent were receiving free or reduced lunch and 8% were receiving special education services. None were eligible to receive services for English Language Learners.

Measures

Early mathematics measures. Three measures were administered to each student: Quantity Discrimination (QD), Missing Number (MN), and Number Identification (NI). Each task was administered individually for one minute, with verbal responses required from the students. The QD task required students to name the larger of two numbers. In the MN task, students were asked to name the missing number in a series of four numbers. The NI task required students to name randomly ordered numbers between 1 and 100. We created alternate forms of the measures constructing each so the task could be administered to both Kindergarten and Grade 1 students. Sample items, descriptive information, and construction procedures for each type of measure are presented in Table 1.

Table 1

Early mathematics measures: Characteristics and item samples

	Quantity Discrimination	Number Identification	Missing Number
Total number of items	63	84	63
Range	Number sets 0-10 and 0-20	Numbers 0-100	Counting by 1's to 20, by 5's to 50, and by 10's to 100
Construction	Randomly select 2 numbers from either set	50% between 0-20, 30% between 0-50, 20% between 0-100	80% counting by 1's; 20% counting by 5's and or?? 10's
Sample Item	13 8	16	6 __ 8 9

Criterion measures. Although our primary purpose was to examine the adequacy of the measures for modeling growth across time, we did examine criterion validity using two additional measures. The criterion measures included teachers' ratings of their students' overall mathematics proficiency, and, for Grade 1 students, a norm-referenced achievement test and a math CBM measure. Teachers used a seven-point Likert scale to rate each student's general proficiency in mathematics relative to other students in the class. Because of the district's assessment program, the students in Grade 1 had taken the *Stanford Early School Achievement Test* (SESAT; Psychological Corporation,1996) prior to the study, so we used each student's standard score on the mathematics subtest as an additional criterion measure. The math CBM score was drawn from the mean of two forms of the Monitoring Basic Skills Progress Computation (Fuchs, Hamlett, & Fuchs, 1998) test, which was administered to participating first grade classes.

Additionally, measures from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) were included as criterion variables. For Kindergarten students, spring scores on the Nonsense Word Fluency measure were gathered. For Grade 1 students, DIBELS Oral Reading Fluency cores from the spring administration were recorded.

Procedures

All data were collected and scored by trained data collectors, who participated in a two hour small-group training session. At the completion of the training, data collectors administered each task to a peer while the trainer observed and completed an 11-item fidelity checklist. All data collectors practiced administration until they were at least 90% accurate.

Interscorer agreement was calculated for approximately 40% of the probes. Agreement was calculated by dividing the smaller score by the larger score and multiplying by 100%. An average agreement level across the selected forms for each measure was calculated. Scoring accuracy was highest for MN (94%), with slightly lower agreement ratios calculated for QD (90%) and NI (88%).

Data were collected at 7 time points during the year, each separated by approximately four weeks (November, December, January, February, early April, late April, and May). Each data collection session lasted approximately 5 to 10 minutes and involved individual administration of the three 1-minute mathematics tasks. During rounds 1, 3, and 7, two forms of each measure were administered to evaluate alternate form reliability. Additionally, round 1 included a pilot involving group administration, with two forms of the early numeracy measures administered to each class as a whole. Round 7 included administration of the DIBELS nonsense word fluency task to the Kindergarten students and the DIBELS oral reading fluency task to the first grade students. The order of the tasks was counterbalanced across students to control for order effects. Students who were absent during a round were given a make-up test if the data could be collected within one week of the original data collection session.

Results

Means and standard deviations for each measure were calculated, and skewness and kurtosis of the distributions examined. Results revealed that the measures met the assumptions of normality for use of Pearson product moment correlations.

Means, standard deviations, skewness, and kurtosis for the seven rounds of Quantity Discrimination measures for Kindergarten in are presented in Table 2.

Table 2

Descriptive Statistics for Quantity Discrimination, Kindergarten Only

Date	Admin.	Measure	Min	Max	M	SD	Skew	Kurtosis
Nov. 05	Indiv.	Form A	0	40	15.88	8.83	.38	-.10
		Form B	0	33	13.27	8.33	.34	-.67
		Average	1.5	35.5	14.58	8.37	.39	-.37
	Group	Form A	0	29	13.38	7.66	-.21	-1.08
		Form B	0	31	14.49	8.17	-.38	-.94
		Average	0.5	28.5	13.94	7.36	-.20	-1.06
Dec. 05	Indiv.	Form A	0	42	18.26	9.36	-.31	-.40
Jan. 06	Indiv.	Form A	1	46	21.60	9.32	-.19	.08
		Form B	0	44	19.40	9.60	-.15	-.20
		Average	0.5	45	20.50	9.24	-.17	.04
Feb. 06	Indiv.	Form A	0	44	22.36	8.98	-.45	.14
Early Apr. 06	Indiv.	Form A	0	42	21.86	9.94	-.38	-.24
Late Apr. 06	Indiv.	Form A	1	40	22.67	8.98	-.29	-.50
May 06	Indiv.	Form A	1	47	27.67	8.83	-.39	.42
		Form B	0	47	25.62	9.54	-.40	.49
		Average	0.5	44	26.65	8.79	-.54	.63

Note. N = 77.

Means, standard deviations, skewness, and kurtosis for the seven rounds of Quantity Discrimination measures for 1st grade are presented in Table 3.

Table 3

Descriptive Statistics for Quantity Discrimination, 1st Grade Only

Date	Admin.	Measure	n	Min	Max	M	SD	Skew	Kurtosis
Nov. 05	Indiv.	Form A	29	17	44	32.10	6.67	-.26	-.51
		Form B	29	16	44	30.10	7.02	-.15	-.52
		Average	29	16.5	42.5	31.10	6.49	-.27	-.35
	Group	Form A	30	14	36	24.33	5.96	.43	-.33
		Form B	30	17	37	23.53	5.38	.79	-.27
		Average	30	15.5	36.5	23.93	5.30	.62	-.10
Dec. 05	Indiv.	Form A	30	15	50	31.07	8.40	-.06	-.19
Jan. 06	Indiv.	Form A	30	18	51	33.93	7.86	-.09	-.07
		Form B	30	17	47	30.23	7.27	.30	-.13
		Average	30	19	49	32.08	7.28	.19	.02
Feb. 06	Indiv.	Form A	30	22	53	36.00	7.47	.43	.02
Early Apr. 06	Indiv.	Form A	30	19	50	32.70	8.31	.35	-.29
Late Apr. 06	Indiv.	Form A	30	13	48	33.63	8.24	-.75	.99
May 06	Indiv.	Form A	30	9	57	37.53	11.07	-.52	.55
		Form B	30	14	55	36.00	10.46	-.11	-.16
		Average	30	18.5	55	36.77	9.92	-.01	-.39
Fall 04			20	5	34	17.70	9.53	.11	-1.02
Spring 05			16	6	37	23.38	7.70	-.90	1.26

Means, standard deviations, skewness, and kurtosis for the seven rounds of Quantity Discrimination measures for both Kindergarten and 1st grade are presented in Table 4.

Table 4

Descriptive Statistics for Quantity Discrimination, Kindergarten and 1st Grade Together

Date	Admin.	Measure	n	Min	Max	M	SD	Skew	Kurtosis
Nov. 05	Indiv.	Form A	106	0	44	20.32	11.00	.17	-.82
		Form B	106	0	44	17.88	10.96	.27	-.83
		Average	106	1.5	42.5	19.10	10.80	.22	-.88
	Group	Form A	107	0	36	16.45	8.73	-.13	-.50
		Form B	107	0	37	17.03	8.51	-.43	-.32
		Average	107	0.5	36.5	16.74	8.18	-.19	-.51
Dec. 05	Indiv.	Form A	107	0	50	21.85	10.75	-.12	-.20
Jan. 06	Indiv.	Form A	107	1	51	25.06	10.50	-.13	-.11
		Form B	107	0	47	22.44	10.22	-.21	-.12
		Average	107	0.5	49	23.75	10.15	-.17	-.04
Feb. 06	Indiv.	Form A	107	0	53	26.19	10.53	-.15	.21
Early Apr. 06	Indiv.	Form A	107	0	50	24.90	10.66	-.25	.08
Late Apr. 06	Indiv.	Form A	107	1	48	25.75	10.04	-.23	-.43
May 06	Indiv.	Form A	107	1	57	30.44	10.45	-.03	.25
		Form B	107	0	55	28.53	10.82	-.06	.34
		Average	107	0.5	55	29.49	10.16	-.05	.51

Means, standard deviations, skewness, and kurtosis for the seven rounds of Missing Number measures for Kindergarten are presented in Table 5.

Table 5

Descriptive Statistics for Missing Number, Kindergarten Only

Date	Admin.	Measure	Min	Max	M	SD	Skew	Kurtosis
Nov. ‘05	Indiv.	Form A	0	22	8.75	5.00	.39	.12
		Form B	0	21	10.12	5.14	-.41	-.26
		Average	0	21.5	9.44	4.75	-.14	.13
	Group	Form A	0	14	6.31	3.77	-.19	-.82
		Form B	0	13	4.90	3.38	.17	-.62
		Average	0	13.5	5.60	3.29	-.12	-.57
Dec. ‘05	Indiv.	Form A	0	24	9.92	4.51	-.25	.71
Jan. ‘06	Indiv.	Form A	0	24	11.30	5.16	-.01	.03
		Form B	0	26	11.09	5.35	-.19	.22
		Average	0	24	11.19	4.95	-.17	.44
Feb. ‘06	Indiv.	Form A	0	31	13.91	5.67	-.18	.47
Early Apr. ‘06	Indiv.	Form A	0	23	12.88	4.99	-.29	-.21
Late Apr. ‘06	Indiv.	Form A	1	25	14.56	5.25	-.16	-.11
May ‘06	Indiv.	Form A	0	24	13.44	4.91	-.18	.45
		Form B	0	26	13.57	4.57	-.40	1.32
		Average	0	25	13.51	4.52	-.31	1.21

Note: n = 77.

Means, standard deviations, skewness, and kurtosis for the seven rounds of Missing Number measures for 1st grade are presented in Table 6.

Table 6

Descriptive Statistics for Missing Number, 1st Grade Only

Date	Admin.	Measure	n	Min	Max	M	SD	Skew	Kurtosis
Nov. 05	Indiv.	Form A	29	4	28	15.38	4.96	.21	.63
		Form B	29	9	25	16.00	3.76	.15	.07
		Average	29	6.5	25	15.69	4.13	.17	.18
	Group	Form A	30	6	22	13.77	3.95	.29	.06
		Form B	30	6	18	11.50	3.54	.15	-1.12
		Average	30	7	20	12.63	3.42	.44	-.57
Dec. 05	Indiv.	Form A	30	8	26	16.13	4.52	.34	-.20
Jan. 06	Indiv.	Form A	30	4	26	15.77	5.22	.09	-.17
		Form B	30	9	26	15.63	3.60	.72	1.56
		Average	30	8	25.5	15.70	3.97	.49	.03
Feb. 06	Indiv.	Form A	30	10	27	19.30	4.16	-.74	.41
Early Apr. 06	Indiv.	Form A	30	11	28	18.80	3.62	.23	.68
Late Apr. 06	Indiv.	Form A	30	11	31	20.47	4.58	.29	-.10
May 06	Indiv.	Form A	30	7	30	17.73	5.64	.13	-.61
		Form B	30	7	27	17.63	4.63	-.50	.17
		Average	30	9.5	28.5	17.68	4.75	.07	-.47
Fall 04			20	1	16	7.65	4.09	.42	-.32
Spring 05			16	1	15	9.56	3.39	-.95	1.65

Means, standard deviations, skewness, and kurtosis for the seven rounds of Missing Number measures for both Kindergarten and 1st grade are presented in Table 7.

Table 7

Descriptive Statistics for Missing Number, Kindergarten and 1st Grade Together

Date	Admin.	Measure	N	Min	Max	M	SD	Skew	Kurtosis
Nov. 05	Indiv.	Form A	106	0	28	10.57	5.78	.33	-.09
		Form B	106	0	25	11.73	5.46	-.39	-.03
		Average	106	0	25	11.15	5.36	-.05	.03
	Group	Form A	107	0	22	8.40	5.08	.31	-.06
		Form B	107	0	18	6.75	4.53	.39	-.37
		Average	107	0	20	7.57	4.58	.38	-.06
Dec. 05	Indiv.	Form A	107	0	26	11.66	5.29	.08	.48
Jan. 06	Indiv.	Form A	107	0	26	12.55	5.53	.07	-.01
		Form B	107	0	26	12.36	5.32	-.32	.45
		Average	107	0	25.5	12.46	5.10	-.17	.43
Feb. 06	Indiv.	Form A	107	0	27	15.42	5.81	-.38	.13
Early Apr. 06	Indiv.	Form A	107	0	28	14.54	5.35	-.32	-.04
Late Apr. 06	Indiv.	Form A	107	1	31	16.22	5.71	-.08	.00
May 06	Indiv.	Form A	107	0	30	14.64	5.45	.11	.27
		Form B	107	0	27	14.71	4.92	-.28	.62
		Average	107	0	28.5	14.68	4.93	-.06	.69

Means, standard deviations, skewness, and kurtosis for the seven rounds of Number Identification measures for Kindergarten are presented in Table 8.

Table 8

Descriptive Statistics for Number Identification, Kindergarten Only

Date	Admin.	Measure	N	Min	Max	M	SD	Skew	Kurtosis
Nov. 05	Indiv.	Form A	77	0	59	18.97	11.16	1.12	1.86
		Form B	77	0	49	14.39	9.75	1.10	1.47
		Average	77	.5	54	16.68	10.21	1.16	1.95
Dec. 05	Indiv.	Form A	77	0	68	24.16	11.76	.54	1.62
Jan. 06	Indiv.	Form A	76	0	69	24.08	12.13	.83	1.83
		Form B	76	0	56	19.99	11.59	.72	.56
		Average	76	1.5	62.5	22.03	11.65	.78	1.23
Feb. 06	Indiv.	Form A	77	2	62	24.16	11.08	.47	1.01
Early Apr. 06	Indiv.	Form A	77	0	57	30.43	12.99	-.42	-.53
Late Apr. 06	Indiv.	Form A	77	2	64	28.57	12.15	.10	.00
May 06	Indiv.	Form A	77	1	73	33.00	12.47	.25	.73
		Form B	77	3	75	28.01	12.03	.87	2.11
		Average	77	4.5	74	30.51	11.60	.65	1.53

Means, standard deviations, skewness, and kurtosis for the seven rounds of Number Identification measures for 1st grade are presented in Table 9.

Table 9

Descriptive Statistics for Number Identification, 1st Grade Only

Date	Admin.	Measure	n	Min	Max	M	SD	Skew	Kurtosis
Nov. 05	Indiv.	Form A	29	21	66	41.48	10.95	.17	-.46
		Form B	29	13	61	37.66	12.79	-.10	-.65
		Average	29	19.5	61.5	39.57	11.18	-.09	-.69
Dec. 05	Indiv.	Form A	30	16	66	45.50	12.95	-.38	-.50
Jan. 06	Indiv.	Form A	30	12	72	41.57	14.28	.22	-.35
		Form B	30	18	75	40.57	14.40	.46	-.18
		Average	30	20.5	73.5	41.07	13.55	.51	-.36
Feb. 06	Indiv.	Form A	30	5	83	44.50	14.96	-.06	1.38
Early Apr. 06	Indiv.	Form A	30	22	74	49.30	12.13	.24	.03
Late Apr. 06	Indiv.	Form A	30	12	78	48.57	14.08	-.31	.68
May 06	Indiv.	Form A	30	25	84	50.17	14.33	.41	.21
		Form B	30	14	77	46.30	14.51	-.06	-.04
		Average	30	19.5	80.5	48.23	14.03	.12	.10
Fall 04			20	3	38	17.10	10.27	.38	-.49
Spring 05			16	8	41	28.00	9.89	-.60	-.38

Means, standard deviations, skewness, and kurtosis for the seven rounds of Number Identification measures for both Kindergarten and 1st grade are presented in Table 10.

Table 10

Descriptive Statistics for Number Identification, Kindergarten & 1st Grade Together

Date	Admin.	Measure	n	Min	Max	M	SD	Skew	Kurtosis
Nov. 05	Indiv.	Form A	106	0	66	25.13	14.96	.64	-.34
		Form B	106	0	61	20.75	14.87	.84	-.12
		Average	106	.5	61.5	22.94	14.63	.71	-.38
Dec. 05	Indiv.	Form A	107	0	66	30.14	15.42	.45	-.22
Jan. 06	Indiv.	Form A	106	0	72	29.03	14.97	.68	.34
		Form B	106	0	75	25.81	15.49	.76	.33
		Average	106	1.5	73.5	27.42	14.89	.71	.33
Feb. 06	Indiv.	Form A	107	2	83	29.86	15.28	.67	.60
Early Apr. 06	Indiv.	Form A	107	0	74	35.72	15.29	-.06	.00
Late Apr. 06	Indiv.	Form A	107	2	78	34.18	15.54	.32	-.09
May 06	Indiv.	Form A	107	1	84	37.81	15.09	.47	.54
		Form B	107	3	77	33.14	15.15	.66	.26
		Average	107	4.5	80.5	35.48	14.64	.62	.38

For Kindergarten students, mean scores on the NI and QD measures were higher than those on MN, with the greatest increase from round 1 to round 7 on the NI measure. Standard deviations for all measures remained relatively consistent over time. There was little evidence of a floor effect, with a maximum of 4 students scoring 0 for any measure in a given round of data collection. For Grade 1 students, mean scores on NI were highest, with QD mean scores next. The greatest increase was again seen for the NI measure. Standard deviations for Grade 1 students for the NI and QD measures were more variable than for Kindergarten students. Grade 1 students never obtained scores of 0 and did not approach a ceiling on any measure.

Descriptives for Criterion Variables

Means, standard deviations, skewness, and kurtosis for criterion variables in Kindergarten only are presented in Table 11.

Table 11

Descriptive Statistics for Criterion Variables, Kindergarten Only

Measure	n	Min	Max	M	SD	Skew	Kurtosis
Teacher Ratings, Winter	76	-2.02	1.79	0	0.97	-.31	-.93
Teacher Ratings, Spring	61	-2.04	1.69	0	0.97	-.44	-.93
Nonsense Word Fluency	77	0	135	25.62	20.53	3.06	-.86

Note. Teacher Ratings were converted to z scores by classroom. Nonsense Word Fluency scores are median of three, items read correctly.

Means, standard deviations, skewness, and kurtosis for criterion variables in 1st grade only are presented in Table 12.

Table 12

Descriptive Statistics for Criterion Variables, 1st Grade Only

Measure	n	Min	Max	M	SD	Skew	Kurtosis
Teacher Ratings, Winter	29	-1.99	1.51	0	.98	-.48	-.80
Teacher Ratings, Spring	30	-1.89	1.13	0	.98	-.45	-1.28
MBSP Computation, Form 1	30	4	16	9.17	3.05	.42	-.40
MBSP Computation, Form 2	30	0	21	9.37	5.33	.08	-.38
MBSP Computation, Average	30	2.5	18.5	9.27	3.81	.53	-.10
Stanford Early Achievement	30	452	607	519.17	37.04	.42	-.16
Oral Reading Fluency	30	6	121	50.50	32.76	.62	-.44

Note. Teacher Ratings were converted to z scores by classroom. MBSP Computation scores are digits correct. Stanford Early Achievement scores are standard scores. Oral Reading Fluency scores are median of three, words read correctly.

Intercorrelations were calculated between each measure presented to students in the study. Correlation coefficients were calculated between the students' average score from two forms of each measure used. Intercorrelations between early math measures for Kindergarten only are presented in Table 13.

Table 13

Intercorrelations Between Early Math Measures, Kindergarten Only

		Quantity Discrim.				Missing Number				Number ID		
		Group	Individual			Group	Individual			Individual		
		F	F	W	S	F	F	W	S	F	W	S
Quantity Discrim.												
Group:	Fall	--										
Individual:	Fall	.47	--									
	Winter	.61	.84	--								
	Spring	.59	.62	.78	--							
Missing Number												
Group:	Fall	.65	.55	.60	.60	--						
Individual:	Fall	.45	.65	.64	.56	.70	--					
	Winter	.46	.64	.74	.74	.74	.79	--				
	Spring	.39	.42	.55	.77	.53	.61	.71	--			
Number ID												
Individual:	Fall	.33	.81	.71	.49	.46	.61	.60	.35	--		
	Winter	.44	.81	.79	.66	.57	.65	.71	.47	.87	--	
	Spring	.40	.61	.69	.76	.44	.53	.63	.58	.56	.70	--

Note. Average score from two forms used for each measure. All correlations significant at $p < .01$.

Intercorrelations between early math measures 1st grade only are presented in Table 14.

Table 14

Intercorrelations Between Early Math Measures, 1st Grade Only

	Quantity Discrim.				Missing Number				Number ID		
	Group	Individual			Group	Individual			Individual		
	F	F	W	S	F	F	W	S	F	W	S
Quantity Discrim.											
Group: Fall	--										
Individual: Fall	.61*	--									
Winter	.34	.70**	--								
Spring	-.05	.55**	.61**	--							
Missing Number											
Group: Fall	.54**	.50**	.30	.18	--						
Individual: Fall	.40*	.64**	.56*	.37	.56**	--					
Winter	.35	.45*	.68**	.22	.47**	.44*	--				
Spring	.28	.57**	.52**	.56**	.66**	.60**	.58**	--			
Number ID											
Individual: Fall	.28	.79**	.71**	.64**	.32	.66**	.48**	.70**	--		
Winter	.08	.56**	.72**	.56**	.17	.41*	.44*	.45*	.80**	--	
Spring	.08	.61**	.60**	.70**	.24	.53**	.41*	.64**	.80**	.73**	--

Note. Average score from two forms used for each measure. * $p < .05$. ** $p < .01$.

Intercorrelations between early math measures for both Kindergarten and 1st grade are presented in Table 15.

Table 15

Intercorrelations Between Early Math Measures, Kindergarten and 1st Grade Together

	Quantity Discrim.				Missing Number				Number ID		
	Group	Individual			Group	Individual			Individual		
	F	F	W	S	F	F	W	S	F	W	S
Quantity Discrim.											
Group: Fall	--										
Individual: Fall	.68	--									
Winter	.68	.86	--								
Spring	.57	.69	.79	--							
Missing Number											
Group: Fall	.76	.76	.68	.61	--						
Individual: Fall	.60	.76	.72	.62	.77	--					
Winter	.55	.68	.78	.67	.72	.77	--				
Spring	.49	.56	.62	.75	.64	.68	.72	--			
Number ID											
Individual: Fall	.57	.89	.79	.65	.70	.75	.65	.56	--		
Winter	.55	.84	.83	.72	.66	.71	.70	.57	.90	--	
Spring	.52	.73	.75	.80	.60	.66	.65	.67	.73	.79	--

Note. Average score from two forms used for each measure. All correlations significant at p < .01.

Intercorrelations between criterion variables teacher ratings for winter and spring and Nonsense Word Fluency for Kindergarten only are presented in Table 16.

Table 16

Intercorrelations Between Criterion Variables, Kindergarten Only

	Teacher Ratings, Winter	Teacher Ratings, Spring
Teacher Ratings, Winter	--	
Teacher Ratings, Spring	.87	--
Nonsense Word Fluency	.45	.42

Note. All correlations significant, $p < .01$. For Winter Teacher Ratings, n = 76. For Spring Teacher Ratings, n = 61, with one classroom not rated. For Nonsense Word Fluency, n = 77.

Intercorrelations between criterion variables teacher ratings for winter and spring, MBSP Computation Average score, Stanford Early Achievement score, and Oral Reading Fluency scores for 1st grade only are presented in Table 17.

Table 17

Intercorrelations Between Criterion Variables, 1st Grade Only

	TR-W	TR-S	MBSP Avg	SEA
Teacher Ratings, Winter (TR-W)	--			
Teacher Ratings, Spring (TR-S)	.85**	--		
MBSP Computation, Average Score (MBSP Avg)	.47*	.58**	--	
Stanford Early Achievement (SEA)	.58**	.58**	.29	--
Oral Reading Fluency (ORF)	.61**	.46**	.23	.25

Note. * $p < .05$. ** $p < .01$. For Winter Teacher Ratings, n = 29; for all other criteria, n = 30.

Reliability

Alternate form reliability correlations are presented in Table 18. Please note that group measures do not appear in Number Identification as no group assessment was conducted using this particular measure. Reliability coefficients were highest for both Kindergarten and Grade 1 students for the QD and NI measures; most were above or approached $r = .80$, which is an acceptable level for educational decision-making. Coefficients were generally lower for the MN measure at both grade levels, ranging from $r = .76$ to .82 for Kindergarten students and from $r =$ .61 to .79 for Grade 1 students. Overall, these reliability coefficients are similar to Lembke and Foegen's (2007) results.

Table 18

Alternate Form Reliability for Early Math Measures

Measure	Grade	Group Admin.	Individual Admin.		
		Fall	Fall	Winter	Spring
Quantity Discrim.	K	.73	.90	.91	.83
	1	.75	.80	.85	.70
	K & 1	.80	.93	.92	.83
Missing Number	K	.69	.76	.77	.82
	1	.67	.79	.61	.71
	K & 1	.82	.82	.77	.81
Number ID	K	--	.91	.93	.79
	1	--	.77	.78	.89
	K & 1	--	.93	.91	.88

Note: All correlations significant at $p < .01$.

Criterion Validity.

Previous research by Lembke and Foegen (2007) examined the validity of these early numeracy measures with criteria such as teacher ratings and group- and individually administered mathematics achievement tests. In the present study, we examined relations between mean scores on two forms of the early mathematics measures and two criteria: teacher ratings of mathematics proficiency and Grade 1 scores on the SESAT (Psychological Corporation, 1996). Previous research has suggested that these validity levels may be further increased by using the average score from two forms (Lembke & Foegen, 2007). Concurrent validity coefficients, were highest for the winter teacher ratings, with moderate to moderately strong values (r = .61 to .75) for both Kindergarten and Grade 1 students. Validity coefficients with the spring teacher ratings were lower for both groups of students, with little or no relationship obtained for a small sample of Grade 1 students. Correlations between Grade 1 students' scores on the early numeracy measures and the SESAT were low to moderate.

Correlation coefficients for Kindergarten were calculated between the early math measures and the two criterion variables, students' scores on Nonsense Word Fluency portion of DIBELS, and teacher ratings for winter and spring. Results for Kindergarten only are presented in Table 19.

Table 19

Criterion Validity of Early Math Measures, Kindergarten Only

		Winter Criterion	Spring Criteria	
		Teacher Ratings	Teacher Ratings	Nonsense Word Fluency
Quantity Discrim.				
Group:	Fall	.36**	.49**	.27*
Individual:	Fall	.60**	.56**	.51**
	Winter	.65**	.59**	.52**
	Spring	.54**	.51**	.46**
Missing Number				
Group:	Fall	.55**	.61**	.41**
Individual:	Fall	.72**	.75**	.57**
	Winter	.69**	.64**	.58**
	Spring	.52**	.49**	.38**
Number ID				
Individual:	Fall	.64**	.64**	.65**
	Winter	.67**	.69**	.69**
	Spring	.54**	.42**	.67**

Note. For Winter Teacher Ratings, n = 76. For Spring Teacher Ratings, n = 61, with one classroom not rated. For Nonsense Word Fluency, n = 77.

Correlation coefficients for 1st grade were calculated between the early math measures and four criterion variables; teacher ratings for winter and spring, Stanford Early Achievement score, MBSP Computation score, and Oral Reading Fluency score from DIBELS. Results for 1st grade only are presented in Table 20.

Table 20

Criterion Validity of Early Math Measures, 1st Grade Only

		Winter Criteria		Spring Criteria		
		Teacher Ratings	Stanford Early Achievement	Teacher Ratings	MBSP Computation	Oral Reading Fluency
Quantity Discrim.						
Group:	Fall	.41*	.53**	.62**	.34	.22
Individual:	Fall	.75**	.50**	.74**	.47*	.58**
	Winter	.73**	.49**	.52**	.39*	.63**
	Spring	.42*	.26	.32	.30	.45*
Missing Number						
Group:	Fall	.48**	.33	.57**	.44*	.30
Individual:	Fall	.53**	.21	.55**	.40*	.51**
	Winter	.58**	.37*	.46*	.19	.55**
	Spring	.56**	.27	.62**	.34	.58**
Number ID						
Individual:	Fall	.78**	.47**	.71**	.34	.81**
	Winter	.73**	.29	.50**	.18	.79**
	Spring	.51**	.13	.30	.04	.70**

Note. * $p < .05$. ** $p < .01$. For Winter Teacher Ratings, n = 29; for all other criteria, n = 30.

Concurrent validity of early math measures were calculated between the early math measures for both grades, and for Kindergarten Nonsense Word Fluency from DIBELS, and for 1st grade, Oral Reading Fluency score from DIBELS. Results are presented in Table 21.

Table 21

Concurrent Validity of Early Math Measures with Reading Criteria

		Kindergarten: Nonsense Word Fluency (n = 77)	1st Grade: Oral Reading Fluency (n = 30)
Quantity Discrimination			
	Round 5	.46	.55
	Round 6	.49	.48
Missing Number			
	Round 5	.50	.67
	Round 6	.37	.58
Number Identification			
	Round 5	.57	.74
	Round 6	.67	.73

Note. All correlations significant, $p < .01$. Early math measures are scores from single forms.

Growth

Two-Level Hierarchical Linear Growth Modeling (HLGM) was used to determine whether Kindergarten and Grade 1 students improved significantly across time with respect to the three math measures: MN, NI, and QD. HLM software version 6.04 (Raudenbush, Bryk, & Congdon, 2007) was used to conduct the analyses. Upon examination of the students' pattern of growth across the seven time points, we determined that some of the growth patterns were not perfectly linear (see Figures 1 – 3 for mean rates at each measurement time for Kindergarten and Grade 1 students on each math measure). Consequently, both linear and polynomial models (using 2nd-, 3rd-, and 4th-order polynomials) were fit to the data in order to represent the appropriate growth pattern for the different math measures.

After fitting linear and different polynomial models to the NI data, we determined that a linear growth model fit the growth pattern for students in both grades (see Figures 1 and 2). Thus, we used the following two-level, unconditional model to assess NI scores across time:

$$\begin{aligned}&Level-1:\\&Y_{ti}=\pi_{0i}+\pi_{1i}Time+e_{ti}\\&Level-2:\\&\pi_{0i}=\beta_{00}+u_{0i}\\&\pi_{1i}=\beta_{10}+u_{1i}\end{aligned},$$

where Y_{ti} is the observed outcome score for student *i* at time *t*; *Time* is the measurement time minus 7; so that π_{0i} is the intercept or the expected outcome score for student *i* at the final measurement time (Round 7); π_{1i} is the slope or the average growth rate in the outcome measure between measurement time 1 and 7 for student *i*; e_{ti} is the error for student *i*; β_{00} is the average intercept or average outcome score for the group of students at the final measurement time; β_{10}

is the average slope or average growth rate between measurement time 1 and 7 for the group of students; and u_{0i} and u_{1i} are the random errors for the average intercept and slope, respectively.

The analyses indicated a significant increase in growth across time on the NI measure for both Kindergarten and Grade 1 students. The average growth rate between each measurement time for Kindergarten and Grade 1 students was 2.04 and 1.42, respectively, so weekly growth rates would be approximately .51 for Kindergarten students and .36 for Grade 1 students on the NI measure. The average intercepts for NI scores were statistically significant for both Kindergarten and Grade 1 students. Further, there were significant differences between students in terms of the final NI scores as well as growth rate of the NI scores within both Kindergarten and Grade 1.

Third-order polynomial models fit the MN and QD data for both Kindergarten and First-grade students (see Figures 2 and 3). Thus, the following two-level, unconditional model was used to assess MN and QD growth across time:

$$\begin{aligned}
&Level-1:\\
&Y_{ti} = \pi_{0i} + \pi_{1i}Time + \pi_{2i}Time^2 + \pi_{3i}Time^3 + e_{ti}\\
&Level-2:\\
&\pi_{0i} = \beta_{00} + u_{0i}\\
&\pi_{1i} = \beta_{10} + u_{1i}\\
&\pi_{2i} = \beta_{20} + u_{2i}\\
&\pi_{3i} = \beta_{30} + u_{3i}
\end{aligned},$$

where Y_{ti} is the observed outcome score for student *i* at time *t*; *Time* is the measurement time minus 7; so that π_{0i} is the intercept or the expected outcome score for student *i* at the final measurement time (Round 7); π_{1i} is the slope or the average growth rate in the outcome measure at time 7 for student *i*; π_{2i} is the rate of acceleration in the outcome measure at time 7 for student

i; π_{3i} is the rate of deceleration of the outcome measure, regardless of time point for student i; e_{ti} is the error for student i; β_{00} is the average intercept or average outcome score for the group of students at the final measurement time; β_{10} is the average slope or average growth rate at time 7 for the group of students; β_{20} is the average slope or average rate of acceleration at time 7 for the group of students; β_{30} is the average slope or average rate of deceleration for the group of students, regardless of time point; and u_{0i}, u_{1i}, u_{2i}, and u_{3i} are the random errors for the average intercept and slopes.

The analyses indicated that all of the intercept and slope coefficients were statistically significant ($p < .05$) for the QD and MN measures. There were significant differences within both Kindergarten and Grade 1 in terms of the final scores. However, the students did not tend to differ significantly on the different growth rate slopes, whether students were increasing or decreasing. Thus, the students within each grade tended to have the same growth rates. Because of the complexity of the models, predicted beginning and end values were computed based on the estimated linear equations (see Figures 1 – 3). The difference between the predicted beginning and end values was then calculated to examine growth from time 1 to time 7. The predicted growth of MN scores from time 1 to time 7 for Kindergarten and Grade 1 students was 5.46 and 2.28, respectively. The predicted growth of QD scores from time 1 to time 7 for Kindergarten and Grade 1 students was 12.24 and 6.36, respectively.

It was of great interest to determine on which measure the students grew most rapidly. To calculate an index (a pseudo effect size measure) of the strength of the growth rates for the different measures, each predicted growth rate from time 1 to time 7 for each measure was divided by its respective standard deviation of final math score. For Kindergarten students, the pseudo effect size index for NI, MN, and QD was 1.09, 1.33, and 1.50, respectively, indicating that Kindergarten students demonstrated the strongest growth rates on the QD measure. The pseudo effect size index for NI, MN, and QD for first grade students was .69, .47, and .66, respectively. Thus, the first grade student's growth rates were the strongest on the NI measure, but the growth rates on the QD measure were fairly comparable.

Discussion and Implications

Previous research in early mathematics has identified measures that are technically adequate for screening students (Clarke & Shinn, 2004; Lembke & Foegen, 2007; Fuchs, et al. 2007). The purpose of this study was to extend the work in early numeracy CBM to examine which measures might model the progress of students across time. It is important to examine the adequacy of measures to be utilized as both measures of performance (for instance, three times per year in screening) and progress (measures given frequently), because some measures might be adequate for use in screening but not capture or model students' progress over time.

We administered three early mathematics measures (NI, QD, and MN) to students in Kindergarten and Grade 1. Previous research using these measures (see Clarke & Shinn, 2004; Lembke & Foegen, 2007) has demonstrated moderate to strong reliability and validity. However, the ability of the measures to reflect progress across time has not been examined.

Based upon the results of this study, it appears that NI, QD, and MN continue to have satisfactory alternate form reliability, with the strongest reliability coefficients (r = .79 to .93) observed for the QD and NI measures for Kindergarten students, and the strongest coefficients for Grade 1 students observed for the QD measure (r = .70 to .85). We examined concurrent criterion validity by correlating scores on the early mathematics CBM measures with Grade 1 standard scores on the SESAT (year) and to teacher rating of mathematics performance. Validity coefficients were highest in the fall across both Kindergarten and Grade 1 when compared to fall teacher ratings (*r* = .55 to .72) and were lowest for Grade 1 students in the spring when compared to teacher ratings, although it should be noted that this was only one teacher's ratings. Validity coefficients with the SESAT were low moderate (*r* = .19 to .46) across the measures. Overall, validity coefficients were lower than those in studies conducted by Clarke and Shinn (2004) and Lembke and Foegen (2007), particularly for the Grade 1 students when examining spring teacher ratings and the SESAT. The lower validity for teacher ratings might be attributed to the small sample size in the spring (one teacher and only 14 students).

The most important question addressed in this study was whether the early mathematics measures could reliably model progress across time and if there were differences between measures in this ability. Results of the HLM analysis indicated that both Kindergarten and Grade 1 students grew significantly on the NI measure linearly over time, with weekly growth rates of approximately .51 for Kindergarten students and .36 for Grade 1 students. These growth rates were greater than those observed in the Lembke & Foegen (2007) research (.17 for Kindergarten and .25 for Grade 1). These findings supported initial work by Chard, Clarke, Baker, Otterstedt, Braun, and Katz (2005) that suggested considerable change from fall to spring on the NI measure, and support the use of this measure to monitor progress over time.

Students' growth was significant from Round 1 to Round 7, but non-linear for the QD and MN measures, so average growth rates could not be calculated. When comparing the strength of growth rates on each of the measures for each grade, Kindergarten students grew at the greatest rate on the QD measure, while Grade 1 students grew at the greatest rate on the NI measure, with QD similar in rate. Both Kindergarten and Grade 1 students grew at the slowest rate on the MN measure. Further research needs to be conducted on these measures to determine if this non-linearity is confirmed, which might indicate less utility as measures of progress.

The HLGM analysis indicated that growth rate slopes for NI were significantly different among students in Kindergarten and Grade 1 respectively, which suggests that NI might serve as a good measure to differentiate students within grade. Visual examination of the graphs in Figures 1 through 3 indicates that there appears to be the greatest differentiation in performance between Kindergarten and Grade 1 students on Number Identification and Quantity Discrimination, while both groups of students performed similarly and had the least growth on the Missing Number measure.

It is important to note that this study was limited by the single geographic location and the lack of diversity in the sample of students, so results should be generalized with these issues as a consideration. Future research should be conducted examining the use of the measures on a weekly basis, with teachers administering and graphing the data to assess effectiveness of instruction. In addition, the non-linearity of the QD and MN measures might be attributed to the reduced reliability of data collection with young children or to form effects. For this reason, the study should be replicated, with a greater number of students, utilizing Item Response Theory analyses to assess reliability of alternate forms.

However, based upon these results, along with previous and present results of reliability and validity studies, it appears that the early numeracy measures of QD, NI, and MN can be utilized as technically adequate measures of both performance (for instance as screening measures) and in the case of NI, progress. This study provides an initial examination of how early numeracy data might be administered and graphed over time to identify students who might not be profiting from instruction, and also can be utilized to determine the effectiveness of interventions for particular students.

References

Baker, S., Gersten, R., Flojo, J. Katz, R. Chard, D. & Clarke, B. (2002). *Preventing mathematics difficulties in young children: Focus on effective screening of early number sense delays* (Tech Rep. No. 0305). Eugene, OR: Pacific Institute for Research.

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 Intervention, 30*(2), 3-14.

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.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30,* 507-524.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36-45.

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

Fuchs, L. S. & Fuchs, D. (2001). Principles for the prevention and intervention of mathematics difficulties. *Learning Disabilities Research and Practice, 16,* 85-95.

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

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

Gersten, R., & Chard, D. (1999). Numbers sense: Rethinking arithmetic instruction for students with mathematical difficulties. *The Journal of Special Education*, *33*, 18-28.

Good, R.H., Gruba, J., & Kaminski, R. (2001). Best practices using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 679–700). Washington, DC: National Association of School Psychologists.

Kroesbergen, E. H. & Van Luit, J. E. H. (2003). Mathematics interventions for children with special education needs: A meta-analysis. *Remedial and Special Education*, *24*, 97-114.

Lembke, E.S., & Foegen, A. (2007). *Identifying indicators of performance in early mathematics for Kindergarten and grade 1 students*. Manuscript submitted for publication.

Miller, S. P., & Hudson, P. J. (2007). Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge, *Learning Disabilities Research and Practice*, *22*, 47-57.

Minskoff, E., & Allsopp, D. (2003). Strategies for academic success: Mathematics. In *Academic success strategies for adolescents with learning disabilities and ADHD* (pp 221-222). Baltimore: Brookes Publishing.

Montague, M. (1996). Assessing mathematical problem solving. *Learning Disabilities Research and Practice*, *11*, 238-248.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

No Child Left Behind Act of 2001, 20 U. S. C. $^{c}{}_{3}$6301 *et seq.*

Perie, M., Grigg, W., & Dion, G. (2005). *The nation's report card: Mathematics 2005.* (NCES 2006-453). U.S. Department of Education, National Center for Education Statistics. Washington, D,C.: U. S. Government Printing Office.

Psychological Corporation. (1996). *Stanford Early School Achievement Test* (9th edition). San Antonio, TX: Harcourt Assessment.

Raudenbush, S., Bryk, A., & Congdon, R. (2007). HLM (Version 6.04) [Computer Software]. Lincolnwood, IL: Scientific Software International, Inc.

Safer, N., & Fleischman, S. (2005). How student progress monitoring improves instruction. *Educational Leadership*, 62(5), pp. 81-83.

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.

Swanson, H. L. (1999). Instructional components that predict treatment outcomes for students with learning disabilities: Support for a combined strategy and direct instruction model. *Learning Disabilities Research and Practice, 14*, 129-140.

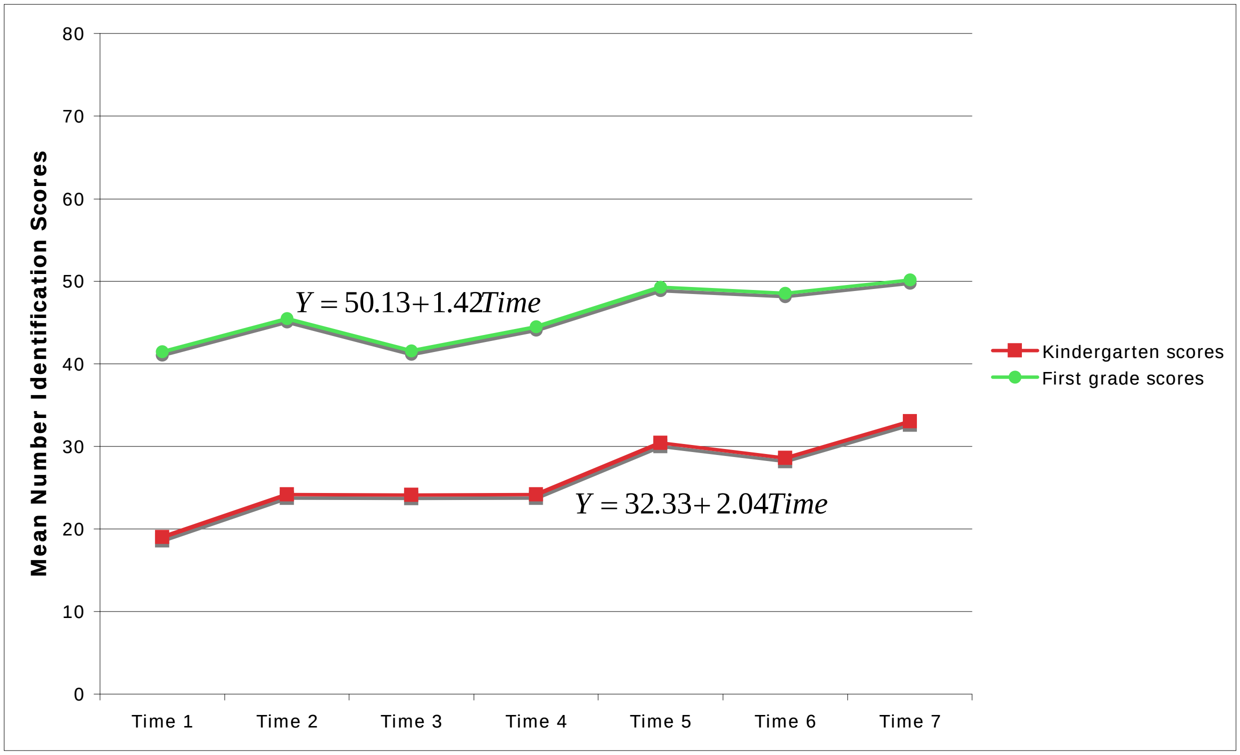
Tindal, G. A., & Parker, R. (1991). Identifying measures for evaluating written expression. *Learning Disabilities Research and Practice,* 6, 211-218.

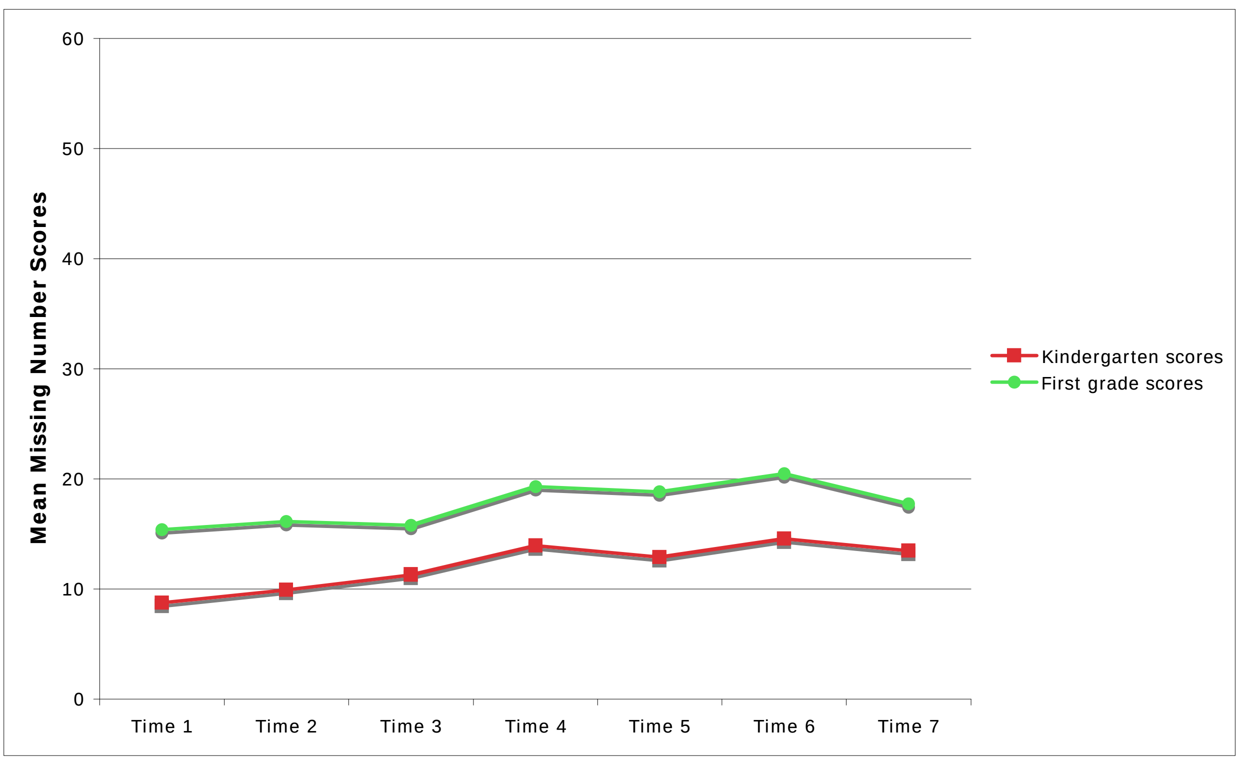
Figure Captions

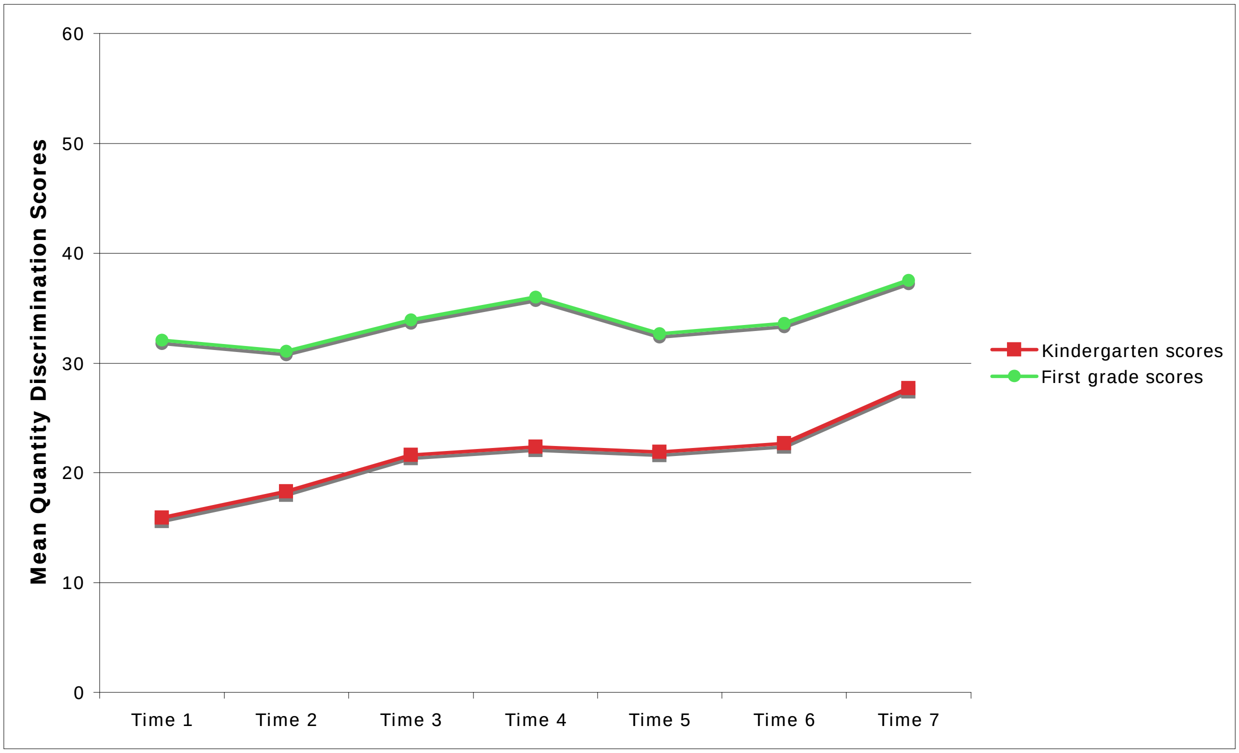
Figure 1. Average Number Identification Growth Trend from Time 1 to Time 7 for Kindergarten and Grade 1 Students.

Figure 2. Average Missing Number Growth Trend from Time 1 to Time 7 for Kindergarten and Grade 1 Students.

Figure 3. Average Quantity Discrimination Growth Trend from Time 1 to Time 7 for Kindergarten and Grade 1 Students.


Mean Number Identification Scores
80
70
60
50
40
30
20
10
0
Time 1
Time 2
Time 3
Time 4
Time 5
Time 6
Time 7
$Y = 50.13 + 1.42 Time$
$Y = 32.33 + 2.04 Time$
Kindergarten scores
First grade scores


Mean Missing Number Scores
60
50
40
30
20
10
0
Time 1
Time 2
Time 3
Time 4
Time 5
Time 6
Time 7
Kindergarten scores
First grade scores


60
50
40
30
20
10
0
Mean Quantity Discrimination Scores
Time 1
Time 2
Time 3
Time 4
Time 5
Time 6
Time 7
Kindergarten scores
First grade scores