TECHNICAL REPORT #6:

Identifying Indicators of Early Mathematics Proficiency in Kindergarten and Grade 1

Erica Lembke and Anne Foegen

August 2005

RIPM Year 2: 2004 – 2005

Dates of Study: October 2004 – May 2005

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.

Identifying Indicators of Early Mathematics Proficiency in Kindergarten and Grade 1

In previous research on the development of curriculum-based measures in math, one of the primary tasks used for screening and progress monitoring was a basic facts probe in which students wrote the answers to single-digit number combinations in single or mixed operations (Marston, 1989). Until recently, there were no measures for students who could not yet compute basic facts, such as students in Kindergarten or Grade 1, or students achieving at that level. With the emphasis on early identification of at-risk students, there is a need to develop measures that can be used to identify students who might be struggling in math as early as possible and to monitor their progress.

In recent years, research on mathematics CBM has been extended to the development of measures for screening and/or progress monitoring students who are not yet able to compute basic facts. In studies by Clarke and Shinn (2004) and Chard, Clarke, Baker, Otterstedt, Braun, and Katz (2005), Kindergarten (n=436) and Grade 1 (n=483) students were administered tasks such as number identification, quantity discrimination, and missing number tasks. In both studies, the researchers varied the way in which students responded (oral or written) and the types of tasks given. At the Kindergarten level, number identification, quantity discrimination, and missing number tasks had the best alternate form reliability and criterion validity, with correlations ranging from .50 to .69 with the criterion, the Number Knowledge Test (Okamoto & Case, 1996). The strongest correlations at the Kindergarten level were found between the missing number tasks and the criterion.

At the Grade 1 level, oral counting, quantity discrimination, number identification, and the missing number tasks all demonstrated adequate reliability and

criterion validity coefficients with several measures (Woodcock Johnson applied problems, Number Knowledge Test, and CBM basic facts probes) ranged from $r = .60$ to $r = .79$.

VanDerHeyden, Witt, Naquin, and Noell (2001) also examined early math measures in a study designed to create a series of reliable and valid CBM probes that could be used as screening tools to identify Kindergarten students in need of academic intervention. For Kindergarten students (n=31) tasks such as circle number, write number, and draw circles had correlations ranging from $r = .44$ to $r = .61$ with the math composite on the Comprehensive Inventory of Basic Skills, Revised (VanDerHeyden et al., 2001). For 4- year olds (n=53) tasks such as choose number, number naming, counting objects, free counting, discrimination, and choosing shapes were correlated with the Brigance Screens (Brigance, 1999) and the Test of Early Mathematics Achievement-2 (TEMA-2; Ginsburg & Baroody, 1990). The strongest correlations were demonstrated with the choosing number and discrimination tasks.

With very little research completed on measures of early math skills, the purpose of our study was to identify reliable and valid curriculum-based measures in early math that could be used with both Kindergarten and Grade 1 students and that could be extended up to higher grade levels. We hoped to replicate and extend the existing work, by introducing new potential measures, investigating different criterion variables, and using the measures with more students. Two studies were completed as part of this first phase of research. Study 1 was completed in Missouri in the fall, with a follow-up to this study completed in Missouri in the spring, and Study 2 was completed in Iowa in the spring.

Study 1 (Missouri)

Method

Participants. One elementary school in a rural town in Missouri participated in Study 1. The district had an elementary school, middle school, and high school within the district. The total enrollment for the district was 1,200 students. Ninety-four percent of the students were Caucasian, 4% African American, 1% Hispanic, and 1% other ethnicities. Thirty-two percent of all students in the district qualified for free or reduced lunch and 11% of students were receiving special education services. Six teachers participated in the study, two in Grade 1 and four in Kindergarten. Eighty-eight students participated in the study, with 56 Kindergarten students and 32 Grade 1 students. Table 1 displays the demographic information for participants in Study 1, fall.

Table 1

Demographic information, Study 1, Fall

	n
Male	48
Female	40
White	83
Other ethnicities	5
Free and Reduced Lunch	23
IEP eligible	8
ELL eligible	1

Early mathematics measures. Four measures were administered to each student: Quantity Discrimination, Quantity Array, Missing Number, and Number Identification. Each measure was administered for one minute. For the Quantity Discrimination task, students were presented with two numbers and asked to name the bigger of two numbers. Each of the three pages of the measure had 21 items arranged in three columns(63 items total). For the Quantity Array measure, students were presented with one or two arrays of dots and asked to name the number of dots in each box. Each of the three pages of the measure had 12 items arranged in two columns (72 items total). For the Missing Number task, students were presented with boxes with a series of four numbers, one of which had been replaced with a blank. Students verbally identified the missing number. Each of the three pages of the measure had 21 items arranged in three columns (63 total items). For the Number Identification task, students were presented with randomly ordered numbers between 1 and 100, each in a separate box, and asked to name each number. Each page of the measure consisted of seven rows of four numbers (81 total items). Three parallel forms of each measure were administered using a one minute duration for each probe. Construction guidelines, administration directions, scoring procedures, and samples for each measure are presented in Appendix A.

Criterion measures. Criterion measures included the teachers' ratings of their students' overall math proficiency, the Woodcock-McGrew-Werder Mini Battery of Achievement (MBA, Woodcock, McGrew, & Werder, 1994), and for the Grade 1 students only, the Stanford Early Achievement Test (1996, Psychological Corporation). For the teacher ratings, each teacher was asked to rate each student's general proficiency in math relative to other students in his/her class, on a Likert scale from 1 to 7, with 1

being least proficient and 7 being most proficient. Teachers were asked to use the entire scale in completing their ratings. The Calculation subtest and Reasoning and Concepts subtest of the Mini Battery of Achievement (Woodcock et al., 1994) were administered to acquire a broad math score for each student. The Grade 1 students had taken the Stanford Early Achievement Test prior to the study, so each student's standard score on the math subtest of the Stanford Early Achievement Test was used as an additional criterion measure.

Procedures. Teachers from the participating district explained the study to students and sent consent letters home. Teachers then collected consent forms and provided students with a pencil as an incentive to return the forms. The first author coordinated the scheduling with the teachers or the building principal at the school, scheduling three data collection sessions for each student.

Students participated in three rounds of data collection, each separated by approximately one week. Rounds 1 and 2 involved administration of four types of math tasks. Three forms of each task were individually administered, with each data collection session lasting approximately 15 minutes per child. Administration of the tasks took place in either a quiet room or hallway. In Round 3, the Mini-Battery of Achievement (Woodcock, McGrew, Werder, 1994) was individually administered to students. The order of the tasks was counterbalanced across students. Students who were absent during the rounds were assessed if they could be assessed within the one-week time limit. If the make-up session could not be completed within the established time limit, that round of data collection was omitted and students were assessed in subsequent rounds of data collection using standard procedures.

All data were collected by trained data collectors. Each data collector for Rounds 1 and 2 participated in small-group training session lasting approximately one hour. Data collectors who participated in Round 3 (MBA) attended an additional training session lasting approximately one hour. Training was provided by the first author. An overview of the study was provided, and then the administration of each task was modeled for the data collectors. The data collectors practiced administering each task and then administered each task to a peer while the trainer observed and completed an 11-item fidelity checklist. Average percentage of accuracy, calculating by dividing the number of items completed correctly by the total number of on the fidelity checklist, was 99% with a range of 91% to 100%.

Scoring and data entry were done by the first author and a research assistant. Ten percent of the probes in each of the two rounds were re-scored to assess inter-scorer agreement; we divided the number of correctly scored papers by the total number of re-scored papers to determine accuracy levels. For round 1 data collection, the scoring accuracy was 94%. For round 2, the scoring accuracy was 96%. The first author checked scoring of 100% of the MBA protocols and found that 97% of the protocols were scored accurately.

Results

For each student, the median of the three scores on each measure was entered into all analyses. Means and standard deviations for each of the measures for each grade are presented in Table 3. The median score of the three forms administered is used in these calculations.

Table 3

Descriptives, First and Second Administration of Measures, Study 1

	Quantity Discrimination		Quantity Array		Number Identification		Missing Number	
	K	1	K	1	K	1	K	1
Range								
▪ 1st admin.	0-33	19-46	1-20	7-28	0-39	17-66	0-13	7-26
▪ *2nd admin.*	*0-35*	*12-52*	*2-19*	*9-28*	*0-38*	*17-57*	*0-17*	*7-30*
Mean								
▪ 1st admin.	13.63	29.73	9.88	15.34	13.02	36.84	6.21	13.72
▪ *2nd admin.*	*16.5*	*32.13*	*10.75*	*15.52*	*15.43*	*37.13*	*7.38*	*14.38*
SD								
▪ 1st admin.	8.22	6.01	3.35	5.52	8.28	9.79	3.39	4.62
▪ *2nd admin.*	*8.51*	*7.77*	*3.75*	*4.83*	*8.66*	*9.7*	*3.74*	*4.58*

Alternate-form and test-retest reliabilities are presented in tables 4 and 5. Alternate-form reliabilities were calculated by correlating scores from the 2nd two forms given during each administration. Because students participated in two rounds of data collection during this first study, two alternate-form reliability coefficients are given for each measure, with the first coefficient calculated using the first round of data and the second coefficient calculated using the second round of data. Test-retest reliabilities were calculated by correlating the average scores from the second two forms administered during the 1st and 2nd administrations. Tests of skewness and Kurtosis were conducted and distributions met the assumptions for use of Pearson product moment correlations.

Table 4

Alternate Form Reliability, First and Second Administrations, Study 1

	Quantity Discrimination		Quantity Array		Number Identification		Missing Number	
	K	1	K	1	K	1	K	1
1st Admin.	.89	.71	.80	.81	.87	.83	.72	.73
2nd Admin.	*.90*	*.86*	*.84*	*.85*	*.91*	*.87*	*.80*	*.75*

Table 5

Test-retest reliability

Quantity Discrimination		Quantity Array		Number Identification		Missing Number	
K	1	K	1	K	1	K	1
.87	.83	.73	.77	.88	.88	.80	.80

Correlation coefficients were calculated between the students' median scores on each of the early math measures at each grade and the two criterion variables, students' standard scores on the Woodcock Johnson Mini Battery of Achievement and the teachers' ratings of students' math proficiency. The correlations with teacher ratings were computed by correlating teacher ratings with each student's average from their median scores in the 1st and 2nd administrations. An additional criterion measure for the Grade 1 students in Study 1 was their standard math scores on the Stanford Tests of Early Achievement. Correlation coefficients from the 1st and 2nd administrations are provided for the MBA and the Stanford Early Achievement Test. Results of the criterion validity analysis are presented in Table 6.

Table 6

Criterion Validity, First and Second Administrations, Study 1

	Quantity Discrimination		Quantity Array		Number Identification		Missing Number	
	K	1	K	1	K	1	K	1
Teacher ratings, (correlated with average scores from 1st and 2nd admin.)	.64***	.58***	.61***	.57***	.65***	.63***	.62***	.63***
MBA								
▪ 1st admin.	.51***	.38*	.36**	.45**	.54***	.47**	.55***	.50**
▪ *2nd admin.*	*.39***	*.31*	*.40***	*.42**	*.51****	*.37**	*.40***	*.44***
Stanford Early Achievement Test								
▪ 1st admin.		.62**		.67**		.54*		.71**
▪ *2nd admin.*		*.62***		*.60***		*.52**		*.56***

*p≤.05, **p≤.01, ***p≤.001

Study 1 Follow-up (Missouri)

Method

Participants. The location for the follow-up study was the same elementary school as study 1, so district demographics will not be repeated. Six teachers participated in the follow-up study, two in Grade 1 and four in Kindergarten. Eighty students participated in the study, with 49 Kindergarten students and 31 Grade 1 students. Table 7 displays the demographic information for participants in the Study 1 follow-up.

Table 7

Demographic Information, Study 1 Follow-up

	n
Male	39
Female	41
White	77
Other ethnicities	3
Free and Reduced Lunch	21
IEP eligible	8
ELL eligible	1

Early math measures. The early math measures that were administered to students were the same as in the fall, with the tasks including Quantity Discrimination, Quantity Array, Missing Number, and Number Identification.

Criterion measures. The criterion measures in the spring included the teacher rating scale (as in the fall); and the Test of Early Mathematics Achievement-3 (Ginsburg & Baroody, 2003), which replaced the Mini Battery of Achievement (Woodcock et al., 1994). In the fall, it appeared that there was a very limited range of scores on the calculation subtest of the MBA, so we decided to utilize a different standardized math test. The TEMA-3 was chosen because of its reliability and validity with younger students.

Procedures. With two exceptions, the procedures for the spring follow-up study were identical to those used in the fall. One change was in the number of data collection

sessions. Because the fall data revealed alternate-form and test-retest reliability estimates that were quite strong, we opted to use a single round of data collection for the early math measures. The second difference was in the criterion measures. Rather than the MBA, we included students' scores on the TEMA as a criterion variable.

Scoring and data entry were done by the first author and a research assistant. Twenty percent of the probes were then re-scored to assess inter-scorer agreement. For Round 1 data collection, the scoring accuracy was 94%. The first author checked scoring of 100% of the TEMA protocols and found a scoring accuracy rate of 79%.

Results

Means and standard deviations for each of the measures for each grade are presented in Table 8. The median score of the three forms administered is used in these calculations.

Table 8

Means and Standard Deviations, Study 1 Follow-up

	Quantity Discrimination		Quantity Array		Number Identification		Missing Number	
	K	1	K	1	K	1	K	1
Range	6-37	21-52	1-24	10-33	8-54	28-76	1-15	5-27
Mean	22.29	34.68	12.32	17.55	27.02	45.77	9.33	14.45
SD	7.07	7.48	3.42	5.01	9.63	12.14	2.74	5.17

Correlation coefficients were calculated between the students' median scores on each of the early math measures at each grade and the two criterion variables, students' standard scores on the TEMA-3 and the teachers' ratings of students' math proficiency. The

correlations with teacher ratings were computed by correlating teacher ratings with each student's median score from each measure. Results of the criterion validity analysis are presented in Table 9. Tests of skewness and Kurtosis were conducted and distributions met the assumptions for use of Pearson product moment correlations.

Table 9

Criterion Validity, Study 1 Follow-up

	Quantity Discrimination		Quantity Array		Number Identification		Missing Number	
	K	1	K	1	K	1	K	1
Teacher ratings	.60***	.57***	.31*	.60***	.57***	.47**	.54***	.61***
TEMA	.49***	.58***	.11	.60***	.36*	.51**	.37**	.58***

*p≤.05, **p≤.01, ***p≤.001

We also examined students' growth from fall to spring across the measures. Median scores on all measures from students that participated in both fall and spring (44 Kindergarten and 28 Grade 1 students) were used to determine growth. Table 10 shows changes in the mean, standard deviation, and range of raw scores from fall to spring.

Table 10

Growth/Change on Measures from Fall to Spring

	Quantity Discrimination		Quantity Array		Number Identification		Missing Number	
	K	1	K	1	K	1	K	1
Range								
▪ Fall, 2nd admin.	2-35	12-52	2-19	9-28	3-38	17-57	1-17	7-30
▪ *Spring*	*6-37*	*24-52*	*7-24*	*10-33*	*8-54*	*28-76*	*1-15*	*5-27*
▪ **Growth**	**-4 to 13**	**-9 to 17**	**-3 to 9**	**-5 to 8**	**-4 to 37**	**-4 to 33**	**-6 to 7**	**-5 to 11**
Mean								
▪ Fall, 2nd admin.	17.61	31.21	11.02	15.52	16.48	36.89	7.82	14.21
▪ *Spring*	*22.61*	*35.14*	*12.72*	*17.57*	*28.02*	*46.71*	*9.41*	*14.75*
▪ **Growth**	**5.00**	**3.93**	**1.70**	**2.05**	**11.55**	**9.82**	**1.59**	**.54**
SD								
▪ Fall, 2nd admin.	7.67	7.86	3.77	4.74	8.74	10.15	3.71	4.87
▪ *Spring*	*6.98*	*7.37*	*3.09*	*5.24*	*9.46*	*12.26*	*2.85*	*5.33*
▪ **Change (fall to spring)**	**4.11**	**5.34**	**2.82**	**3.69**	**7.81**	**9.29**	**2.86**	**3.65**

Study 2 (Iowa)

Method

Participants. Two districts participated in Study 2. District 1 was in a rural community, only a few miles from an urban community. The district had four schools within the district: a preschool to second grade building, a third to fifth grade building, a middle school, and a high school. The total enrollment for the district was 2,243 students, with 52% of the enrollment population being male, 96% white, 2% Hispanic, and 1.8% other ethnicities. Just over ten percent of the students qualified for free and reduced lunch, and 15.5 % of students were receiving special education services. Four teachers were invited by the building principal to participate in the study (two kindergarten

teachers and two Grade 1 teachers) because they were piloting a new math curriculum that school year. Seventy-five students participated in the study (34 Kindergarten students and 41 Grade 1 students). Table 11 displays the demographic characteristics of the student participants from District 1.

District 2 was a suburban community and included four schools: a preschool to third grade building, a fourth/fifth grade building, a middle school, and a high school. The total enrollment for the district was 1,424 students, with 53% being male, 93% white, 3.6 % Hispanic, and 3% other ethnicities. Nearly 42% of the students qualified for free and reduced lunch, and 16.7% of students were receiving special education services. All four kindergarten and four Grade 1 teachers in the building were invited and consented to participate in the study. One hundred forty nine students participated in the study (63 K and 86 Grade 1 students). Table 11 displays the demographic characteristics of the student participants in Study 2.

Table 11

Demographics, Study 2

	District 1	District 2	Total
Male	43	63	106
Female	32	86	118
White	75	136	211
Hispanic	0	7	7
Other ethnicities	0	4 (2 missing)	4 (2 missing)
Free and Reduced Lunch	4	61	65
IEP eligible	0	1	1
Total # participants	75	149	224

Measures. The early math measures used in Study 2 were identical to those in Study 1 with one exception. The students in Study 2 did not complete the number identification task. The criterion measures for Study 2 included teacher ratings and student scores on the Mini-Battery of Achievement.

Procedures. Teachers from the participating districts explained the study to students and sent consent letters home. Teachers then collected consent forms and provided students with a university logo pencil as an incentive to return the forms. A graduate research assistant (GRA) coordinated the scheduling with the teachers or the building principal at each school, scheduling three data collection sessions for each student.

Students participated in three rounds of data collection; each separated by approximately one week (no less than five days, no more than eight days). Rounds 1 and 2 involved three types of early math tasks. Three forms of each task were individually administered, with each data collection session lasting approximately 15 minutes per child. Administration of the tasks took place in either a quiet room or hallway. In Round 3, the Mini-Battery of Achievement was individually administered to students. The order of the early math tasks was counterbalanced across students to control for any possible order effects. Spanish administration of all 3 rounds was provided for three students identified by teachers as needing administration in their native language. Students who were absent during one of the rounds were assessed if data collection could be completed within the one-week time limit. If the make-up session could not be completed within the established time limit, that round of data collection was omitted and students were assessed in subsequent rounds of data collection using standard procedures.

All data were collected by trained data collectors. Each data collector for Rounds 1 and 2 participated in small-group training session lasting approximately one hour. Data collectors who participated in Round 3 (MBA) attended an additional training session lasting approximately one hour. Training was provided by the GRA or the second author. An overview of the study was provided, and then the administration of each task was modeled for the data collectors. The data collectors practiced administering each of the tasks, then administered each task to a peer while the trainer observed and completed an 11-item fidelity checklist. Overall, the average percentage of accuracy on the fidelity checklist was 99%, with a range of 91-100%. The only item on the checklist that data

collectors missed was marking a line under the last number the student said prior to the end of the one-minute time period.

Scoring and data entry were done by both by the GRA and an hourly worker. Fifteen percent of the probes in each of the two rounds were then re-scored to assess inter-scorer agreement. Point-by-point agreement was used to calculate an accuracy percentage for each probe. The means and ranges for scoring accuracy are displayed in Table 12 by probe type. The GRA checked scoring of 15% of the MBA protocols. These results are also displayed in Table 2.

Table 12

Interscorer Agreement for Early Numeracy Tasks

Task	Average % of accuracy	Range
Missing Number	99.7%	86.7-100%
Quantity Array	99.7%	86.7-100%
Quantity Discrimination	99.2%	86.4-100%
MBA	100%	

Results

Means and standard deviations for each of the measures for each grade are presented in Table 13. The median score of the three forms administered is used in these calculations.

Table 13

Demographics, Study 2

	Quantity Discrimination		Quantity Array		Missing Number	
	K	1	K	1	K	1
Range						
▪ 1st admin.	1-39	6-54	6-28	3-37	0-20	1-28
▪ *2nd admin.*	*1-46*	*16-63*	*2-24*	*5-40*	*0-24*	*2-33*
Mean						
▪ 1st admin.	18.83	31.63	12.85	18	9.44	13.90
▪ *2nd admin.*	*20.63*	*37.02*	*10.47*	*18.35*	*9.49*	*15.12*
SD						
▪ 1st admin.	9.51	8.27	3.82	5.75	3.88	4.67
▪ *2nd admin.*	*9.61*	*9.16*	*3.81*	*7.01*	*4.65*	*5.10*

Alternate-form and test-retest reliabilities are presented in tables 14 and 15. Alternate-form reliabilities were calculated by correlating number correct scores from the 2nd two forms given during each administration. Test-retest reliabilities were calculated by correlating the average scores from the second two forms administered during the 1st and 2nd administrations. Tests of skewness and Kurtosis were conducted and distributions met the assumptions for use of Pearson product moment correlations.

Table 14

Alternate Form Reliability, First and Second Administrations, Study 2

	Quantity Discrimination		Quantity Array		Missing Number	
	K	1	K	1	K	1
1st Admin.	.88	.89	.74	.83	.75	.80
2nd Admin.	*.91*	*.84*	*.79*	*.85*	*.80*	*.83*

Table 15

Test-retest Reliability, First and Second Administrations, Study 2

Quantity Discrimination		Quantity Array		Missing Number	
K	1	K	1	K	1
.84	.91	.81	.85	.84	.89

Correlation coefficients were calculated between the early math measures at each grade and the two criterion variables, students' standard scores on the Woodcock Johnson Mini Battery of Achievement and the teachers' ratings of students' math proficiency. The correlations with teacher ratings were computed by correlating teacher ratings with each student's average from their median scores in the 1st and 2nd administrations. Results from the 1st administration, with the mean score utilized, are presented. Results of the criterion validity analysis are presented in Table 16.

Table 16

Criterion Validity Coefficients, Study 2

	Quantity Discrimination		Quantity Array		Missing Number	
	K	1	K	1	K	1
Teacher ratings, (correlated with average scores from 1st and 2nd admin.)	.65***	.69***	.45***	.52***	.61***	.64***
MBA ▪ 1st admin. ▪ *2nd admin.*	.49*** *.64***	.45*** *.49***	.48*** *.45***	.37*** *.41***	.43*** *.58***	.46*** *.44***

*p≤.05, **p≤.01, ***p≤.001

Discussion/Implications

Conclusions from Study 1 and Study 1 Follow-up:

Ranges of scores for both Kindergarten and Grade 1 students indicated that few students received a score of zero. During the first phase of Study 1, there was 1 Kindergarten student who scored 0 on several of the measures and 4 other students who scored 0 on the missing number task during the 1st administration. However, no student scored 0 on any of the measures during the follow-up in the spring. Students also demonstrated growth on all measures from fall to spring, with the most growth demonstrated for the number identification measure for both K and Grade 1 students. However, the range of scores for number identification indicated that students, and particularly Grade 1 students, were also approaching a ceiling for this measure. If students do not have room to continue to grow on the measure, the measure is not a viable option for progress monitoring. The greatest standard deviations were seen in the Quantity Discrimination and number identification measures, indicating that students varied more widely in their performance on these measures. This variation is important, as we are seeking to identify measures that might differentiate among students.

With respect to reliability, for Kindergarten students, the obtained pattern of correlations was consistently stronger for the Quantity Discrimination and number identification measures for alternate form reliability (.89-.90 for Quantity Discrimination and .87-.91 for number identification), although Quantity Array and missing number had reliabilities that were moderately strong (.80-.84 for Quantity Array and .72-.80 for missing number). For Grade 1 students, Quantity Array and Number Identification had the strongest alternate form reliabilities (.81-.85 for Quantity Array and .83-.87 for

Number Identification, while Quantity Discrimination also demonstrated a strong coefficient in the second administration, .86.

For test-retest reliability, Quantity Discrimination and Number Identification were the strongest measures for Kindergarten students, with coefficients ranging from .87 (Quantity Discrimination) to .88 (Number Identification). Missing Number also showed promise, with a correlation of .80. For Grade 1 students, results for test-retest reliability were parallel to that of K students, with Quantity Discrimination and Number Identification demonstrating the strongest reliabilities (Quantity Discrimination, .83, and Number Identification, .88) and Missing Number showing promise (.80).

With respect to criterion validity in the fall of Study 1, correlations between teacher ratings and the various measures ranged from .57 to .64 for the 1st administration, and .31 to .61 for the 2nd administration. For Kindergarten students, all measures correlated above .60 with teacher's ratings. For Grade 1 students, no particular measure stood out, with correlation coefficients ranging from .57 to .63 across the four measures. In the follow-up phase of Study 1, average teacher ratings were once again correlated with students' scores on the measures and for K students, Quantity Discrimination (.60) and Number Identification (.57) had the strongest correlation coefficients. For Grade 1 students, Quantity Array (.60), Missing Number (.61), and Quantity Discrimination (.57) had the strongest coefficients.

When examining correlations between the standardized test, the Mini Battery of Achievement, and students' scores on the measures, the strongest measures for K students appeared to be Quantity Discrimination (.51), Number Identification (.54), and Missing Number (.55). For Grade 1 students, the strongest correlation coefficients were

with the Quantity Array (.45), Number Identification (.47), and Missing Number (.50) measures. An additional analysis was conducted for Grade 1 students, correlating the four measures with scores on an additional standardized test, the Stanford Early Achievement Test. We obtained the strongest correlations for Quantity Discrimination (.62), Quantity Array (.67) and Missing Number (.71), indicating that it appears there are a variety of measures that might be used. In the follow-up study in the spring, students' scores were correlated with an additional standardized measure, the Test of Early Math Achievement (TEMA-3). Coefficients between Kindergarten students' scores and the TEMA ranged from .11 (Quantity Array) to .49 (Quantity Discrimination), with the strongest correlation for Quantity Discrimination. For Grade 1 students, coefficients ranged from .51(Number Identification) to .60 (Quantity Array), with the strongest correlations for Quantity Discrimination (.58), Quantity Array (.60), and Missing Number (.58).

When examining growth in the Study 1 follow-up, both Kindergarten and Grade 1 students showed the most growth when measured using the Quantity Discrimination and Number Identification measures. Kindergarten students had a mean growth rate of 5 problems correct from fall to spring on the Quantity Discrimination measure and mean growth rate of 11.54 problems correct on Number Identification. Grade 1 students had a mean growth rate of 3.93 problems correct on the Quantity Discrimination measure and 9.82 problems correct on the Number Identification measure. When examining the standard deviation of the change in means from fall to spring, it appears that most Kindergarten and Grade 1 students made growth on the Number Identification measure.

Conclusions from Study 2

For Study 2, one Kindergarten student scored 0 when Missing Number was administered the first time, and three Kindergarten students scored 0 when Missing Number was administered the second time. The means for each measure indicated that both Kindergarten and Grade 1 students grew from 1st to 2nd administration, with the exception of the Kindergarten students on the Quantity Array measure. Some Grade 1 students were approaching the ceiling on the Quantity Discrimination measure. Both Kindergarten and Grade 1 students tended to complete the most items on the Quantity Discrimination measure (means of 18.83 and 20.63 for Kindergarten and 31.63 and 37.02 for 1st). Standard deviations were also greatest for both Kindergarten and Grade 1 students for Quantity Discrimination, indicating that this measure had the greatest variation in student scores.

With respect to reliability, for Kindergarten students, Quantity Discrimination demonstrated the strongest alternate-form reliability (.88-.90), although Missing Number had a reliability coefficient that was moderately strong (.80). For Grade 1 students, Quantity Discrimination and Quantity Array had the strongest alternate-form reliabilities (.84-.89 for Quantity Discrimination and .83-.85 for Quantity Array), while Missing Number also demonstrated strong coefficients .80-.83.

For test-retest reliability, Quantity Discrimination and Missing Number were the strongest measures for Kindergarten students, with coefficients at .84. Quantity Array also showed promise (.81). For Grade 1 students, results for test-retest reliability were parallel to that of Kindergarten students, with Quantity Discrimination and Missing

Number demonstrating the strongest reliabilities (Quantity Discrimination, .91, and Missing Number, .89) and Quantity Array showing promise (.85).

With respect to criterion validity, correlations between teacher ratings and the various measures ranged from .45 to .69. For Kindergarten students, the correlation coefficient with Quantity Discrimination was .65 and with Missing Number was .61. A similar pattern of results was observed for Grade 1 students, with the correlation coefficient with Quantity Discrimination at .69 and with Missing Number at .64.

When examining correlations between the standardized test, the Mini Battery of Achievement, and students' scores on the measures, the strongest measures for Kindergarten students appeared to be Quantity Discrimination (.64) and Missing Number (.58). For Grade 1 students, the strongest correlations were realized with the Quantity Discrimination (.49) and Missing Number (.46) measures. An abbreviated listing of primary results is included in Appendix B.

For Kindergarten students, Quantity Discrimination and Number Identification have the strongest reliabilities, and Quantity Discrimination appears to have the strongest validity, with Missing Number and Number Identification close behind. For Grade 1 students, Quantity Discrimination and Number Identification have the strongest reliabilities, with Missing Number demonstrating the strongest validity and Quantity Discrimination the next strongest although students appear to be approaching a ceiling on Number Identification. The Quantity Array measure did not demonstrate strong reliability and validity results for Kindergarten students nor, in most cases, for Grade 1 students. Future research should examine the utility of using the promising measures early math measures to demonstrate growth over time.

References

Brigance, A. (1999). *Comprehensive Inventory of Basic Skills* (rev. ed.). North Billerica, MA: Curriculum Associates.

Chard, D. J., 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.

Ginsburg, H. P., & Baroody, A.J. (1990). *Test of Early Mathematics Ability, 2nd edition.* Austin, TX: Pro-Ed, Incorporated.

Ginsburg, H. P. & Baroody, A.J. (2003). *Test of Early Mathematics Ability, 3rd edition.* Austin, TX: Pro-Ed, Incorporated.

Marston, D. (1989). Curriculum based measurement: What it is and why do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford.

Okamoto, Y., & Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. In R. Case & Y. Okamoto (Eds.), *The role of central conceptual structures in the development of children's thought: Monographs of the Society for Research in Child Development* (Vol. 1-2, pp. 27-58). Malden, MA: Blackwell Publishers.

Stanford Early Achievement Test (9th edition, 1996). Psychological Corporation. Harcourt Publishing.

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

Woodcock, R. W., McGrew, K. S., & Werder, J. K. (1994). *Woodcock-McGrew-Werder Mini-Battery of Achievement*. Itasca, IL: Riverside Publishing.

Appendix A

Early Math Measures

Missing Number

Construction:

1) Use forward counting sequence
2) 70% of the problems are counting by 1's and 30% are counting by 2's, 5's, and 10's
3) Problem type is randomly selected
4) Number that the problem starts with is randomly selected (0-7 for count by 1's, 2-14 for count by 2's, 5-35 for count by 5's, and 10-70 for count by 10's)
5) Includes problems counting by 1-digit from 0-10, by 2's to 20, by 5's to 50, and by 10's to 100.
6) For counting by 1-digit, the blank varies
7) For counting by 2's, 5's, and 10's, the blank is at the end.
8) 3 numbers are given, with a blank in-between or at then end (student completes the pattern by stating the 4th number)

Directions for Missing Number:

1. Place the student copy in front of the student.

2. Place the examiner copy on a clipboard and position so the student cannot see what the examiner records.

3. Say these specific directions to the student:

> **"Look at the paper in front of you. Each box has three numbers and a blank."** (Point to the first box). **"What number goes in the blank?"**

4. Correct Response:

> **"Good. The number is 3**." (Point to the second box.)

Incorrect Response:

> **"The number that goes in the blank is 3**. **You should have said 3 because 3 comes after 2 (0, 1, 2, 3)."** (Point to the second box.)

5. Say to the student:

"Here is another example. What number goes in the blank?"

6. Correct Response:

"Good. The number is 2." (Point to the third box).

Incorrect Response:

"The number that goes in the blank is 2. You should have said 2 because 2 comes after 1 (1, 2, 3, 4)." (Point to the third box.)

7. Say to the student:

"Here is another example. What number is goes in the blank?"

8. Correct Response:

"Good. The number is 20." (Turn the page).

Incorrect Response:

"The number that goes in the blank is 20. You should have said 20 because 20 comes after 15 when you are counting by 5's (5, 10, 15, 20)." (Turn the page.)

9. Say to the student:

"The paper in front of you has boxes with three numbers and a blank in each of them. When I say begin, I want you to tell me what number goes in the blank in each box. Start here and go across the page (demonstrate by pointing). **Try each one. If you come to one that you don't know, I'll tell you to try the next one. Are there any questions? Put your finger on the first one. Ready, begin."**

10. Start your stopwatch. If the student fails to attempt (does not give the answer to the first problem) after 3 seconds, tell the student to

"Try the next one."

11. For at least the first 2 to 3 rows of problems, you may need to prompt the student by pointing to the next box and saying

"Tell me the number that goes in the blank."

12. On the administrator copy, write the number that the student says in the blank next to each problem number.

13. The maximum time for each item is 3 seconds. If a student does not provide an answer within 3 seconds, tell the student to

"Try the next one."

14. If the student comes to the end of the page, turn the page to the next page of problems.

15. At the end of 1 minute, draw a line under the last item completed and say

"Stop."

Scoring Rules

Rule 1: If a student correctly identifies the number score the item as correct.

Rule 2: If the student states any number other than the item number score the item as incorrect.

Rule 3: If a student hesitates or struggles with a problem for 3 seconds tell the student to "try the
next one" and score the item as incorrect.

Rule 4: If a student skips a problem, score the problem as incorrect.

Rule 5: If a student skips an entire row, mark each problem in the row as incorrect.

Missing Number task (actual task is three pages long)

Example 1

0 1 2 ___	1 ___ 3 4	5 10 15 ___

0 ___ 2 3	7 8 9 ___	___ 1 2 3
4 ___ 6 7	5 10 15 ___	8 10 12 ___
2 3 4 ___	___ 5 6 7	7 8 ___ 10
4 ___ 6 7	3 4 5 ___	___ 7 8 9
14 16 18 ___	1 2 3 ___	30 40 50 ___
___ 5 6 7	5 ___ 7 8	70 80 90 ___
7 8 ___ 10	4 6 8 ___	___ 1 2 3

Number Identification

Construction:

1) Random numbers 0-100
2) 50% of the numbers are between 0-20, 33% are between 0-50, and 17% are between 0-100.

Directions for Number Identification:

1. Place the student copy in front of the student.

2. Place the examiner copy on a clipboard and position so the student cannot see what the examiner records.

3. Say these specific directions to the student:

> **"Look at the paper in front of you. There are numbers in boxes."** (Point to the first box). **"What number is this?"**

4. Correct Response:

> **"Good. The number is 6**." (Point to the second box.)

Incorrect Response:

> **"This number is 6."**(Point to the second box.)

5. Say to the student:

> **"Here is another example. What number is this?"**

6. Correct Response:

> **"Good. The number is 15."** (Point to the third box).

Incorrect Response:

> **"This number is 15."** (Point to the third box.)

7. Say to the student:

> **"Here is another example. What number is this?"**

8. Correct Response:

> **"Good. The number is 1."** (Turn the page).

Incorrect Response:

> **"This number is 1."** (Turn the page.)

9. Say to the student:

> **"The paper in front of you has boxes with numbers in them. When I say begin, I want you to tell me what number is in each box. Start here and go across the page** (demonstrate by pointing). **Try each one. If you come to one that you don't know, I'll tell you to try the next one. Are there any questions? Put your finger on the first one. Ready, begin."**

10. Start your stopwatch. If the student fails to attempt (does not give the answer to the first problem) after 3 seconds, tell the student to

> **"Try the next one."**

12. For at least the first 2 to 3 rows of problems, you may need to prompt the student by pointing to the next box and saying

> **"What number is this?"**

12. On the administrator copy, write the number that the student says in the blank next to each problem number.

13. The maximum time for each item is 3 seconds. If a student does not provide an answer within 3 seconds, tell the student to

> **"Try the next one."**

14. If the student comes to the end of the page, turn the page to the next page of problems.

15. At the end of 1 minute, draw a line under the last item completed and say

> **"Stop."**

Scoring Rules

Rule 1: If a student correctly identifies the number score the item as correct.

Rule 2: If the student states any number other than the item number score the item as incorrect.

Rule 3: If a student hesitates or struggles with a problem for 3 seconds tell the student to "try the
next one" and score the item as incorrect.

Rule 4: If a student skips a problem, score the problem as incorrect.

Rule 5: If a student skips an entire row, mark each problem in the row as incorrect.

Number identification task (actual task is 3 pages long):

Example

6	15	1	44

Number Identification, page 1—Student copy

6	26	39	9
16	5	18	8
6	8	4	0
18	30	16	2
18	94	17	22
7	64	47	9
1	34	24	97

Quantity array

Construction:

- Generated random numbers from 1-10
- Generated random numbers from 1-2 to determine whether there would be a single or double set for each problem
- Eliminated every other "2" in the list, so there are more 1's (single sets) than 2's (double sets)
- Square up sets
- Maximum for each set in a line will be 5
- If there are 2 sets with objects in line, the lines need to be vertical.
- In the finished probes, make sure there is a range of problem types (some in lines, some in sets, more single sets than double sets.)

Administration directions

1. Place the student copy in front of the student.

2. Place the examiner copy on a clipboard and position so the student cannot see what the examiner records.

3. Say these specific directions to the student:

> **"Look at the paper in front of you. There are 2 boxes with dots in them.** (Point to the first box). **How many dots are in this box?"**

4. Correct Response:

> **"Good. The number is 3**." (Point to the second box.)

Incorrect Response:

> **"The answer is 3** (demonstrate by counting the 3 dots), **because there are three dots in the box. Let's try another one."** (Point to the second box.)

5. Say to the student:

> **"Here is another example. How many dots are in this box?"**

6. Correct Response:

> **"Good. The number is 4."** (Turn the page).

Incorrect Response:

> **"The answer is 4** (demonstrate by counting the 4 dots), **because there are four dots in the box."** (Turn the page).

7. Say to the student:

> **"The paper in front of you has boxes with dots in them. When I say begin, I want you to tell me how many dots are in each box. Start here and go across the page** (demonstrate by pointing). **Try each one. If you come to one that you don't know, I'll tell you to try the next one. Are there any questions? Put your finger on the first one. Ready, begin."**

8. Start your stopwatch. If the student fails to attempt (does not start counting the dots or does not give the answer to the first problem) after 5 seconds, tell the student to

> **"Try the next one."**

9. For at least the first 2 to 3 rows of problems, you may need to prompt the student by pointing to the next box and saying

> **"Tell me how many dots are in the box."**

10. On the administrator copy, write the number that the student says in the blank next to each problem number.

11. The maximum time for each item is 5 seconds. If a student does not provide an answer within 5 seconds, tell the student to

> **"Try the next one."**

12. If the student comes to the end of the page, turn the page to the next page of problems.

13. At the end of 1 minute, draw a line under the last item completed and say

> **"Stop."**

Scoring Rules

Rule 1: If a student correctly identifies the number score the item as correct.

Rule 2: If the student states any number other than the item number score the item as incorrect.

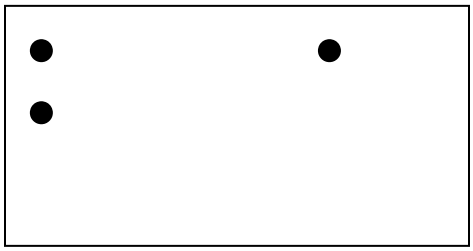
Rule 3: If a student hesitates or struggles with a problem for 5 seconds tell the student to "try the
next one" and score the item as incorrect.

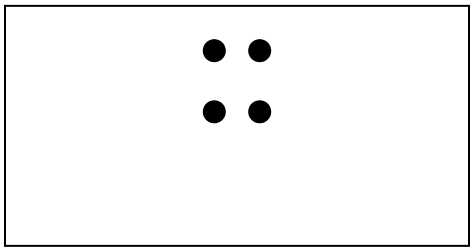
Rule 4: If a student skips a problem, score the problem as incorrect.

Rule 5: If a student skips an entire row, mark each problem in the row as incorrect.

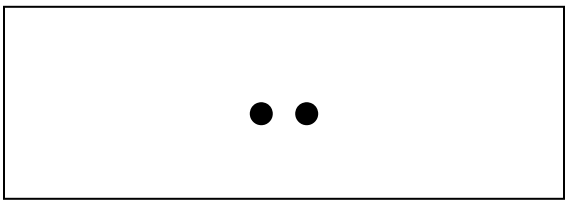
Quantity Array Task (actual task is 3 pages long)

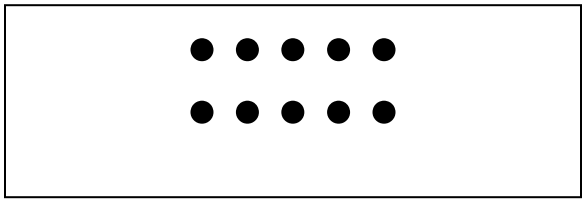
Example

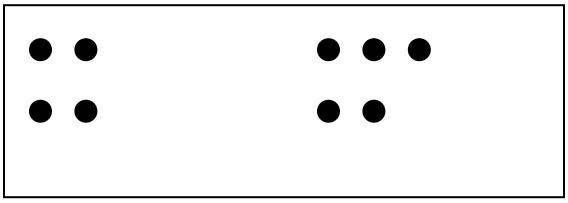


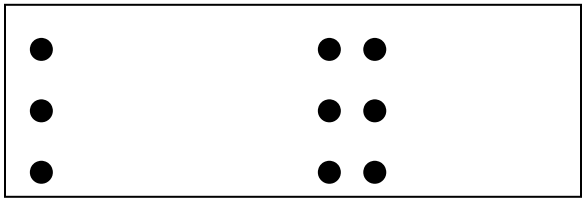


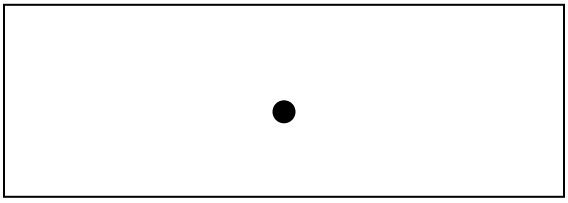
Quantity Arrays, page 1—student copy

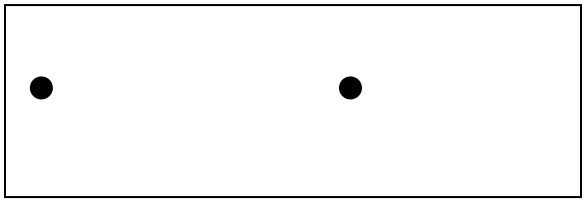


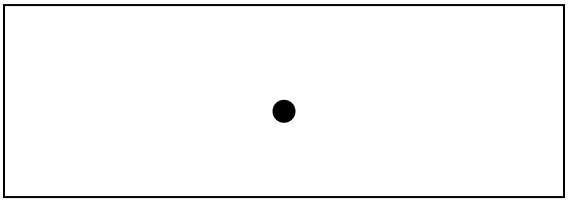


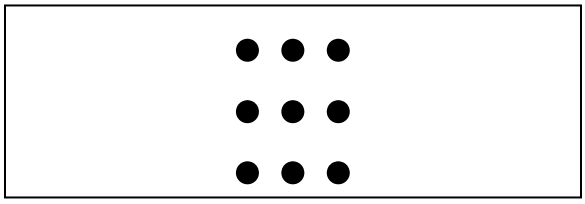


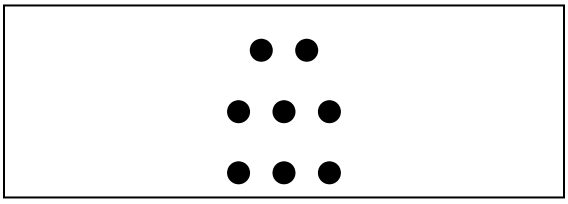


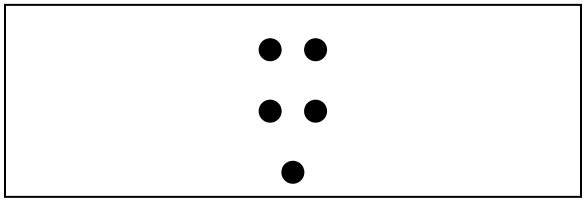


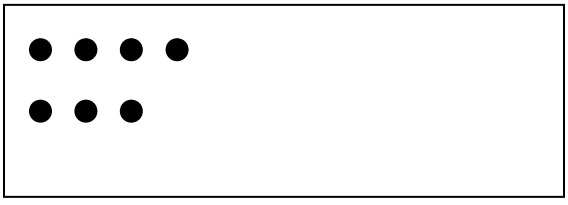












Quantity Discrimination

Construction:

- Used number sets 0-10 and 0-20
- Randomly selected either 0-10 or 0-20
- From the selected number set, selected two numbers for each problem
- If the next two random numbers are identical (i.e., next two numbers are both 2's) eliminate one, and move to the next number.

Directions for Quantity Discrimination:

1. Place the student copy in front of the student.

2. Place the examiner copy on a clipboard and position so the student cannot see what the examiner records.

3. Say these specific directions to the student:

> **"Look at the paper in front of you. In each row there are some boxes with numbers in them."** (Point to the first set of boxes in the top row). **" I want you to tell me the number that is bigger."**

4. Correct Response:

> **"Good. 7 is bigger than 1.**" (Point to the second set of boxes in the top row.)

Incorrect Response:

> **"The number that is bigger is 7. You should have said 7 because 7 is bigger than 1."** (Point to the second set of boxes in the top row.)

5. Say to the student:

> **"Here is another example. Tell me the number that is bigger."**

6. Correct Response:

> **"Good. 6 is bigger than 2."** (Point to the third set of boxes in the top row).

Incorrect Response:

> **"The number that is bigger is 6. You should have said 6 because 6 is bigger than 2."** (Point to the third set of boxes in the top row.)

5. Say to the student:

> **"Here is another example. Tell me the number that is bigger."**

6. Correct Response:

> **"Good. 8 is bigger than 0."** (Turn the page).

Incorrect Response:

> **"The number that is bigger is 8. You should have said 8 because 8 is bigger than 0."** (Turn the page.)

7. Say to the student:

> **"The paper in front of you has boxes with two numbers in each box. When I say begin, I want you to tell me which number is bigger. Start here and go across the page** (demonstrate by pointing). **Try each one. If you come to one that you don't know, I'll tell you to try the next one. Are there any questions? Put your finger on the first one. Ready, begin."**

8. Start your stopwatch. If the student fails to attempt (does not give the answer to the first problem) after 3 seconds, tell the student to

> **"Try the next one."**

10. For at least the first 2 to 3 rows of problems, you may need to prompt the student by pointing to the next box and saying

> **"Tell me which number is bigger."**

10. On the administrator copy, write the number that the student says in the blank next to each problem number.

11. The maximum time for each item is 3 seconds. If a student does not provide an answer within 3 seconds, tell the student to

"Try the next one."

12. If the student comes to the end of the page, turn the page to the next page of problems.

13. At the end of 1 minute, draw a line under the last item completed and say

"Stop."

Scoring Rules

Rule 1: If a student correctly identifies the number score the item as correct.

Rule 2: If the student states any number other than the item number score the item as incorrect.

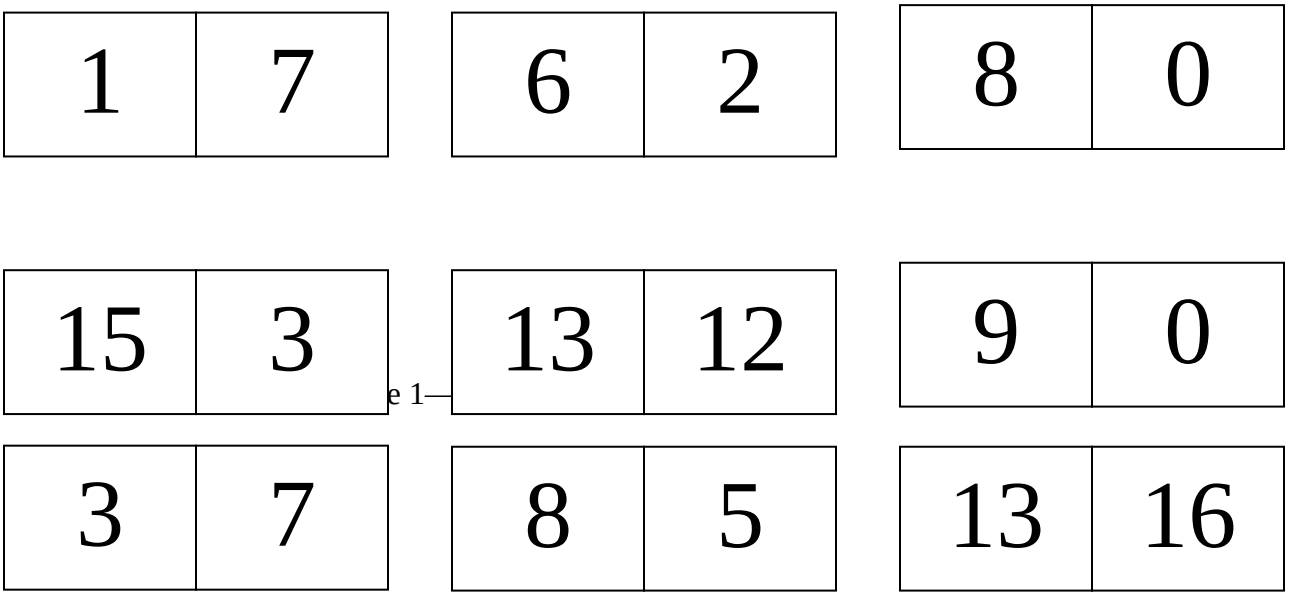
Rule 3: If a student hesitates or struggles with a problem for 3 seconds tell the student to "try the
next one" and score the item as incorrect.

Rule 4: If a student skips a problem, score the problem as incorrect.

Rule 5: If a student skips an entire row, mark each problem in the row as incorrect.

Quantity Discrimination Task (actual task is 3 pages long)

Example 1



e 1—

16	2

13	12

9	0

4	11

8	1

1	11

5	0

2	10

10	9

7	1

8	7

16	2

0	7

1	0

9	1

6	0

9	19

5	1

Appendix B

Results Summary

Reliability (Missouri):

- K—QD, NI (alt. form)
- K-QD, NI (test-retest)
- 1st—NI (alt. form)
- 1st- NI (test-retest)

Reliability (Iowa)

- K—QD (alt. form)
- K-QD, MN, QA (test-retest)
- 1st—QD (alt. form)
- 1st-QD, MN(test-retest)

Validity (Missouri)

- K—QD, NI, MN (teacher ratings)
- K—QD, NI, MN (MBA)
- K—QD (TEMA)
- 1st—all measures similar (teacher ratings)
- 1st—QA, NI, MN (MBA)
- 1st—QD, QA (SEA)
- 1st—QD, QA, MN (TEMA)

Validity (Iowa)

- K—QD, MN (teacher ratings)
- K—QD, QA (MBA)
- 1st—QD, MN (teacher ratings)
- 1st—QD, MN (MBA)