

COVER PAGE

Exploring the Promise of a Number Line Assessment to Help Identify Students At-Risk in Mathematics

Abstract This manuscript presents the results from a study to investigate the technical characteristics of two versions of a number line assessment (NLA 0–20 and NLA 0–100). The sample consisted of 60 kindergarten and 46 first grade students. Both number line versions had sufficient alternate form and test–retest reliability. The NLA 0–20 had low and the NLA 0–100 had low to moderate correlations with math achievement. Results indicated that the NLA 0–100 explained a small, but unique portion of the variance in first grade mathematics performance when controlling for performance on the Assessing Student Proficiency in Early Number Sense (ASPENS) a set of early numeracy screening measures. We discuss study results related to the utility of adding number line assessment tasks to mathematics screening batteries and propose additional areas of research.

Published Assessment for Effective Intervention advanced online publication DOI 10.1177/1534508418791738

Authors

Ben Clarke, PhD, Mari G. Strand Cary, PhD, Lina Shanley, PhD, Marah Sutherland, BA

Center on Teaching and Learning, University of Oregon, 1600 Millrace Dr., Suite 207, Eugene, OR 97403, USA

University of Oregon, 1600 Millrace Dr., Suite 207, Eugene, OR 97403, USA

Funding Source Number R324A110286 (IES), H327S140019 (Office of Special Education and Rehabilitative Services), R305A170044 (IES)

Title of Grants KinderTEK: Teaching Early Knowledge of Whole Number Concepts Through Technology, Stepping Up KinderTEK: Expanding Early Math Instruction for Students with Disabilities, Evaluation of the KinderTEK iPad Math Program

Exploring the Promise of a Number Line Assessment to Help Identify Students At-Risk in Mathematics

Assessment for Effective Intervention
1–10

© Hammill Institute on Disabilities 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1534508418791738

aei.sagepub.com



Ben Clarke, PhD¹, Mari G. Strand Cary, PhD¹, Lina Shanley, PhD¹,
and Marah Sutherland, BA¹

Abstract

This manuscript presents the results from a study to investigate the technical characteristics of two versions of a number line assessment (NLA 0–20 and NLA 0–100). The sample consisted of 60 kindergarten and 46 first grade students. Both number line versions had sufficient alternate form and test–retest reliability. The NLA 0–20 had low and the NLA 0–100 had low to moderate correlations with math achievement. Results indicated that the NLA 0–100 explained a small, but unique portion of the variance in first grade mathematics performance when controlling for performance on the Assessing Student Proficiency in Early Number Sense (ASPENS) a set of early numeracy screening measures. We discuss study results related to the utility of adding number line assessment tasks to mathematics screening batteries and propose additional areas of research.

Keywords

number sense, mathematics, screening

Over the past two decades, there has been significant interest in improving the performance of our nation's children in the area of mathematics (National Mathematics Advisory Panel, 2008; National Research Council, 2001). Interest driven, in part, by low levels of mathematics performance. Recent National Assessment of Educational Progress (National Assessment of Education Progress (NAEP) data indicate overall low levels of performance at fourth grade with only 40% of students being classified as at or above proficiency. Findings are of greater concern for students from minority populations, low socioeconomic backgrounds, English learners, and students with disabilities with a range of 14% to 26% being classified as at or above proficiency. NAEP longitudinal trends (2015) also indicate that performance levels have remained relatively stagnant since 2007 after a prolonged period of sustained improvement. To address long standing concerns with mathematics achievement, a focus on early intervention seems warranted given findings from multiple longitudinal studies (Duncan et al., 2007; Morgan, Farkas, & Wu, 2009; Morgan, Hillemeier, Farkas, & Maczuga, 2014). Investigations of mathematics development have demonstrated strong relationships between early mathematics risk status at kindergarten and fifth grade (Morgan et al., 2009) and documented that students who exited kindergarten at-risk in mathematics were 17 times more likely to have

continuous and persistent difficulties in mathematics through late elementary and middle school (Morgan et al., 2014). Thus, in the absence of targeted efforts, it is likely that students with early difficulties in mathematics will continue to struggle as they encounter more advanced mathematics (Jordan, Glutting, & Ramineni, 2010).

Despite recognition of the importance of mathematical knowledge and its acquisition as a fundamental goal of schooling, systematic efforts to increase mathematics achievement are limited. One proposed framework for increasing student achievement across academic areas is the use of Multitier Systems of Support (MTSS) or Response to Intervention (RTI) service delivery models with a focus on the prevention of academic difficulties (Lembke, McMaster, & Stecker, 2010). However, a recent analysis found that while over 70% of schools reported the use of MTSS or RTI frameworks to support the acquisition of early literacy skills only 35% reported similar support in mathematics (Balu et al., 2015). Results in this vein are not surprising given the persistent lack of time spent on mathematics (La Paro et al.,

¹University of Oregon, Eugene, USA

Corresponding Author:

Ben Clarke, University of Oregon, 1600 Millrace Dr., Suite 207, Eugene, OR 97403, USA.

Email: clarkeb@uoregon.edu

2009), the systematic focus on reading instruction as the primary charge in the early elementary grades (Clarke, Doabler, & Nelson 2014), and the complexity of translating research into practice for complex RTI procedures in mathematics such as data-based individualization to intensity instruction (Schumacher, Edmonds, & Arden, 2017).

Efforts to develop parallel systems in mathematics are dependent upon advances in key components of RTI systems (Gersten et al., 2009) and while system-wide supports targeting mathematics achievement in school are not as common place as in reading, within the last decade, researchers have made significant advances in two key areas: screening for risk status and the delivery of evidence-based interventions (Fuchs, Fuchs, & Compton, 2012). There have several intervention programs in mathematics developed and validated for use in the early elementary grades (e.g., Clarke et al., 2014, 2016; Dyson, Jordan, & Glutting, 2013; Fuchs et al., 2005; Sood & Jitendra, 2013) that, as called for by experts, target the development of number sense and whole number understanding (Author et al., 2009; Frye et al., 2013). In parallel, researchers have also focused on building corresponding screening systems to identify students in need of additional support (see Fuchs et al., 2007, and Gersten et al., 2012, for comprehensive reviews). The purpose of this study was to investigate potential new approaches to screening for risk status in early mathematics.

Typically, screening measures in early mathematics are developed with curriculum-based measurement (CBM) design parameters (Deno, 1985) that emphasize measures not only demonstrate strong psychometric properties including the capacity to model student growth but that they are simple, efficient, and easily understood. Measures are designed as general outcome measures with a focus on assessing a student's overall understanding of mathematics. Measures commonly used in early mathematics screening batteries can be considered to fall into one of two camps. The first set assesses readiness skills such as engaging in rote counting or identifying numerals. The second set is more focused on assessing key mathematics concepts (Clarke, Gersten, Dimino, & Rolfhus, 2011). For example, measures focused on student understanding of magnitude (i.e., comparing two numerals and identifying the lesser or greater magnitude) and the ability to engage in strategic counting (i.e., identifying the missing numeral from a sequence of numerals) have consistently shown promise for identifying students at-risk in mathematics (Author et al., 2012; Foegen, Jiban, & Deno, 2007; Fuchs et al., 2007).

The relative success of magnitude comparison and strategic counting as screeners may be due to their ability to tap into the development of a mental number line (Berch, 2005). Number line development enables children to engage in a range of mathematics tasks as they link their informal understanding of number and beginning number sense to

the formal system of numbers taught in schools (Gersten & Chard, 1999). For example, students may use a mental number line when applying counting strategies to solve addition and subtraction problems. The use of a mental number line is also critical to tasks related to understanding numerical magnitudes which encapsulates a student's ability to "comprehend, estimate, and compare the sizes of numbers" (Fazio, Bailey, Thompson, & Siegler, 2014). Researchers have theorized that the mental number line operates as the primary central conceptual structure (Case et al., 1996) by which student's organize and integrate new information into their understanding of number systems (Siegler, Thompson, & Schneider, 2011) making number line performance a potential valuable mechanism by which to identify at-risk status.

Several researchers (e.g., Booth & Siegler, 2006; Geary, Hoard, Nugent, & Byrd-Craven, 2008; Laski & Siegler, 2007) have attempted to assess the development of a mental number line and its role in understanding magnitude through a number line estimation task. This task requires students to map a given numerical value (typically an Arabic numeral, but also nonsymbolic magnitude representations) onto a horizontal number line. Traditionally, the number line is labeled with two defined end points, and the student is asked to indicate where a value lies between these points (e.g., placing 5 on a number line with endpoints of 0 and 20). Estimation error is calculated by comparing the individual's placement of the numerical value to the actual location of the target value on the number line.

More accurate placement of numbers on the number line estimation task is related to higher math achievement, greater proficiency in solving arithmetic problems, and better performance on magnitude comparison tasks (Booth & Siegler, 2006; Fazio et al., 2014; Laski & Siegler, 2007). In addition, typically achieving students perform better on the number line estimation task compared to students with mathematics learning disabilities (Geary et al., 2008). As a student progresses in their schooling, they continue to reference a mental number line as they move onto more challenging tasks such as comparing the magnitude of fractions (Siegler et al., 2011). Jordan and colleagues (2013) found that even when accounting for variables such as fact fluency or working memory, third grade number line task performance was the stronger predictor of fourth grade mathematics achievement including fraction understanding. Such results support the theory that number line tasks tap into a broader understanding of number and that number line tasks can be used to detect difficulties as students transition from working with the whole number system to the rational number system (Siegler et al., 2011). The relation between student performance on number line estimation tasks and mathematics achievement points to its potential utility as a screening tool for students with mathematics difficulties. By assessing students' mental number line in a more direct

manner that differs from current mathematics screening practice, we may, increase the probability of accurately detecting early mathematics difficulties.

Purpose and Research Questions

Our orientation to this study was based on the field's current practice of using CBM to efficiently screen for mathematics risk status thus we chose to approach the work by investigating whether number line estimation measures added value to the screening process when added to a set of commonly used early numeracy CBM screeners. To date, no studies have looked at the inclusion of a number line estimation task within a standard CBM early numeracy screening battery. Due to increased interest in investigating constructs associated with magnitude, new formats to investigate those constructs, and interest in improving early numeracy screening systems (Author et al., 2011b), this study specifically focused on answering two questions related to the potential use of two iPad number line assessments (NLAs), one spanning 0–20 (NLA 0–20) and one spanning 0–100 (NLA 0–100):

1. What are the psychometric properties, including alternate form and test–retest reliability and concurrent and predictive validity by kindergarten and first grade, of two iPad administered number line assessments (NLA 0–20 and NLA 0–100)?
2. To what extent does performance on a number line assessment (NLA 0–20 and NLA 0–100) add incremental validity to a battery of early numeracy curriculum based measures (ASPENS) in predicting student mathematics achievement (EasyCBM)?

Method

Participants

The study was conducted in a mid-sized public school district in the Pacific Northwest during a 5-week, district-sponsored summer school program. The district enrolls 1,817 K-3 students. Of those, 77% are economically disadvantaged, 18% have identified disabilities, and 10% are English learners. Within-year mobility is 14.4%. The major of students are white (67%), with 21% Hispanic/Latino, 7% multiracial, 2% Black/African American, 2% Asian, 1% American Indian/Alaska Native, and less than 1% Native Hawaiian/Pacific Islander. The program operated 2 hr per day, 4 days per week and provided free breakfast and lunch to participants. It served exiting kindergarten and first grade students determined by the district to be at-risk in reading. At-risk status was determined for kindergarten students if they fell below the 30th percentile on an EasyCBM Word Reading Fluency measure and for first grade students if they fell below the 20th percentile on an

EasyCBM Passage Reading Fluency measure. Reading instruction was the major content focus however the district also provided math instruction to all students for 30 min per day through the use of an individualized computer delivered program, NumberShire, focused on review and practice with whole number concepts. Those students who had attended kindergarten and first grade the previous year ($n = 134$ “outgoing” or “exiting” students) were invited to participate in the study. A number of families opted out ($n = 14$) or were not present during testing ($n = 14$), thus our sample included 106 students (60 exiting first graders, 46 exiting kindergarteners).

Design

We examined within- and across-student performance on assessments administered by the research team at the beginning and end of summer school and standardized measures administered by the district (i.e., at the end of the previous school year and beginning of the following school year).

Measures

We administered two iPad adaptations of the number line estimation task (*NLA 0–20* and *NLA 0–100*) and two curriculum-based assessments (*Assessing Student Proficiency in Early Number Sense* and *EasyCBM*). Accommodations were not provided as part of the administration procedures.

NLA 0–20 and NLA 0–100. The general number line task first appeared for such uses in the literature in 2003 (Siegler & Opfer), but has been adapted for different ages, purposes, and presentation formats (Strand Cary, Laski, Shanley, & Clarke, 2014). For this study, we developed the NLA 0–20 task and NLA 0–100 task (administered through iPads and modeled after a 26-item paper/pencil number line task by Laski (2013). During the task, students were presented with a horizontal line on the iPad screen while the iPad app (i.e., a female voice) explained the concept of the number line and that one end represents 0 and the other represents 20 (in the case of the NLA 0–20) or 100 (in the case of the NLA 0–100). Students practiced placing these numbers and received affirmative verbal feedback (e.g., “Right! Zero goes here on the number line.”) or corrective verbal and visual feedback (e.g., “That spot belongs to another number. Zero goes HERE on the number line”; “Close, but that spot belongs to another number. Twenty goes HERE on the number line.”). The spot was marked by the number presented sliding to the appropriate spot on the number line.

Students then placed randomly presented target numbers on the number line. Students' responses and the time it took them to hit the “submit” button were logged by the app. The app also calculated item-specific errors (e.g., if

the number was 4 and the student placed it at 10.25, the error would be +6.25; if the student placed it at 1.73, the error would be -2.27). The resulting data file included information about students' performance on the sample items as well as each target number, associated error, and submission time, as well as cumulative and mean absolute error (i.e., the main outcome variable). Minimum scores for both NLAs were 0 and maximum absolute scores were 15 (mean) and 240 (cumulative) for the NLA 20 and 77.58 (mean) and 2017 (cumulative). Responses were not flagged for any particular response pattern (e.g., quick and random). Overall test time (approximately 3.5 min for the NL 0–20 and 5 min for the NL 0–100 task). During the NLA 0–20 task, students were presented with most numbers between 1 and 19. To avoid giving students clear “anchors” during the task, 5, 10, and 15 were not presented. For purposes of this study, two forms of the NLA 0–20 were utilized. The first form (20a) presented 2, 4, 7, 8, 11, 13, 16, and 19 randomly in the first half of the assessment, then 1, 3, 6, 9, 12, 14, 17, and 18 randomly in the second half. The second form (20b) reversed the halves. During the NLA 0–100 task, 26 numbers (the same as those used in Laski, 2013) were presented.

Assessing Student Proficiency in Early Number Sense (ASPENS; Sopris). The first grade ASPENS assessment consists of three, 1- to 2-min timed, individually administered measures that assess the ability to compare two numerals and determine which is greater (Magnitude Comparison) identify the missing numeral in a string of three numerals (Missing Number) and solve simple addition and subtraction computation problems that cross 10 (Basic Arithmetic Facts and Base 10; Clarke et al., 2011). Subtest scores are calculated and weighted to form an overall ASPENS composite score. We administered the first grade winter version at the beginning of summer school and the spring version at the end. The authors report test-retest reliability ranging from .71 to .90. Concurrent and predictive validity with the TerraNova Third Edition is reported as ranging from .57 to .63 and as .63 respectively.

EasyCBM. EasyCBM is a mathematics measure that emphasizes conceptual understanding over basic computation and is based on the Common Core State Standards for Mathematics (Anderson, Alonzo, & Tindal, 2010; Common Core State Standards Initiative, 2010). Each individualized math assessment is computer-administered and contains 30 (kindergarten) or 35 (first grade) items. For first grade, the measures exhibit strong internal consistency (Cronbach's alpha from .78 to .89) and concurrent validity (correlation of .73 with the TerraNova; Anderson et al., 2010). The EasyCBM system generates percentile scores and raw scores. We used percentile scores to provide additional contextual information regarding the study sample.

Assessment Procedures

All procedures were approved by the participating district and the University's institutional review board (IRB). Parent information letters with opt-out postcards were mailed 2 weeks before the study start date and student assent was procured during the initial test session. The week before pretesting, a nine-person data collection team comprising five seasoned university assessors and four new-hires was trained to administer the ASPENS and number line tasks. The 2-hr training included training in administration logistics (e.g., counterbalancing, technical troubleshooting), practice assessments and in-training reliability (i.e., fidelity of administration) for the ASPENS. A standard of 90% was required to be considered reliable, but retraining to 100% reliability for all data collectors took place in the days following the training.

Initial assessments were administered within the first week of summer school. The project coordinator (i.e., a veteran data collector) shadow scored assessors' first ASPENS administrations and conducted informal, in-field observations to verify in-field reliability of administration for both the ASPENS and number line tasks. Students first completed ASPENS, then both number line assessments; the entire testing battery took 15 to 20 min per student. Random assignment was used to determine whether students first completed the NLA 0–20 or NLA 0–100, as well as which of the two NLA 0–20 forms they completed (i.e., 20a or 20b). The same data collection team administered final assessments a few days before the end of summer school. Half of the students started with ASPENS and the other half started with NLA tasks. At posttest, students completed the NLA 0–20 and NLA 0–100 in the opposite order (and form) from their initial assessment (i.e., if they completed pretesting in the order NLA 0–100/NLA 0–20b, post testing would be NLS 0–20a / NLA 0–100). ASPENS tests were scanned and scored by Teleform. Number line assessments were scored and saved to .csv files by the app itself. The district provided spring EasyCBM scores (of the school year preceding the study) and fall EasyCBM scores (of the school year following the study) in a .csv file. EasyCBM is used by the district beginning in Grade 1 (thus pretest data is not available for our exiting kindergarten sample).

Statistical Analyses

Univariate descriptive analyses were performed on all measures of number line knowledge and mathematics achievement. Pearson's r correlation coefficients were used to examine the relationships among the study variables. Reliability coefficients and relevant correlations were examined to evaluate the psychometric properties of the NLA 0–20 and NLA 0–100. Then, hierarchical multiple linear regression models were generated to address our second

research question. Namely, we tested whether NLA 0–100 scores predicted EasyCBM percentile rank scores above and beyond ASPENS composite scores for outgoing kindergarten and first grade students. The NLA 0–20 scores were not tested due to low correlations with other math measures used in the study. The hierarchical multiple linear regression models analyzed here included two steps. In the first step, EasyCBM percentile rank scores were regressed on ASPENS composite scores. Then, the NLA 0–100 scores were added to the model to determine the extent to which NLA 0–100 scores were able to explain additional variance in mathematics achievement percentile rank. In the model results, we reported the R^2_{total} and R^2_{change} statistics to describe the proportion of variance in the dependent variable captured by the independent variables in each block. Basic descriptive analyses were conducted using SPSS 20.0 for Mac OS (IBM Corp, 2011) and all subsequent models were investigated using the maximum likelihood estimation in Mplus 7.1 (Muthén & Muthén, 2013).

Results

Descriptive statistics and correlations between the study variables are displayed in Table 1. Across both grade level samples, correlations between all variables ranged from .01 to .94. As expected, all of the mathematics screening subtests (ASPENS) were positively correlated with one another in both the kindergarten and first grade samples. Greater range in ASPENS subtest correlations were observed in the kindergarten sample, $r = .21-.85$, as compared to the first grade sample, $r = .61-.83$. Whereas the NLA 0–20 scores demonstrated few statistically significant correlations with other measures, the NLA 0–100 was positively associated with a number of other measures in both samples, $r = .01-.48$.

Reliability

Test–retest reliability for the NLA 0–100 was measured at .72 and .70 in kindergarten and first grade, respectively. Alternate form reliability was assessed for the NLA 0–20 and range from .57 – .61. Given questions about the accuracy of Cohen’s alpha for accurately measuring internal consistency (Sijtsma, 2009), λ_2 values were estimated in addition to Cohen’s alpha at each administration occasion. Across samples, NLA 0–100 internal consistency ranged from $\alpha = .88-.90$, $\lambda_2 = .88-.90$; NLA 0–20 internal consistency ranged from $\alpha = .83-.93$, $\lambda_2 = .84-.94$.

Incremental Validity

Together, ASPENS composite scores and the NLA 0–100 performance explained 17% of the variance in mathematics achievement percentile rank for the kindergarten sample,

$F(2, 35) = 3.48, p < .05$; however, neither measure was statistically significantly associated with mathematics percentile rank in the kindergarten sample. Although not statistically significant, the NLA 0–100 explained an additional 7% of the variance in mathematics percentile rank. Results of the hierarchical linear regression analyses for first grade (see Table 2) indicated that the addition of the NLA 0–100 significantly improved the prediction of mathematics achievement percentile rank. ASPENS composite scores, and NLA 0–100 performance were statistically significantly associated with mathematics percentile rank in the first grade sample, $F(2, 37) = 17.29, p < .001$. Together the measures explained 49% of the variance in mathematics achievement percentile rank for the first grade sample, and NLA 0–100 scores uniquely explained 13% of the variance in the first grade sample, $\Delta F(1, 36) = 8.47, p < .05$.

Discussion

Our work in investigating the NLA was guided by an orientation toward the use of CBM like measures that requires considerations related to efficiency of use for screening, the capacity to model growth for progress monitoring, and a reflection of current best practices in early mathematics screening (Author et al., 2012). Because these considerations guided our thinking, we investigated the utility of the NLA when added to a standard early numeracy screening battery. Thus, rather than studying the technical characteristics of the NLA in isolation, we examined if the inclusion of the NLA would explain additional variance in math achievement to such an extent that the expenditure of additional resources and time to collect number line data were justified. Results indicate low to moderate (Salvia & Ysseldyke, 2004) alternate form and test–retest reliability and stronger internal consistency reliability. Validity correlations were highly variable with stronger results at first grade compared to kindergarten and the NLA 0–100 measure demonstrating more statistically significant correlations. In addition, the NLA 0–100 measure explained an additional 7% of the variance on the kindergarten criterion measure and a statistically significant 13% of additional variance on the first grade criterion measure. Given the additional time required to administer the NLA measures (approx. 3.5–5 min) relative to the additional variance explained in student mathematics performance and the need for caution in interpreting results of correlational studies (Thompson, Diamond, McWilliam, Snyder, & Snyder, 2005), the results from this study do not support a change in current practice but they do raise several interesting questions and directions for future research.

Approaches to validating assessments typically include initially working with a sample of students across the skill spectrum. A limitation of this study was that the sample was a sample of convenience and consisted of students who

Table 1. Descriptive Statistics and Correlations for All Study Variables for Kindergarten ($n = 46$) and First Grade ($n = 60$) Samples.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	M (SD)
1. Magnitude comp. T1													3.89(6.45) 14.47(6.88)
2. Missing number T1	.85 .79												2.18(3.24) 9.05(4.53)
3. Base 10 T1	.21 .67	.35 .61											0.45 (1.13) 4.28 (3.92)
4. ASPENS comp. T1	.95 .92	.95 .89	.44 .86										7.73 (11.29) 34.88 (18.08)
5. 20 Number line T1	.36 .08	.44 .18	.23 .09	.42 .14									4.89 (3.14) 3.94 (2.83)
6. 100 Number line T1	.31 .29	.40 .40	.12 .23	.36 .35	.71 .63								29.30 (9.42) 25.05(11.94)
7. Magnitude comp. T2	.90 .86	.79 .79	.23 .65	.87 .85	.45 .09	.35 .30							4.84 (6.47) 15.10 (7.12)
8. Missing number T2	.81 .82	.88 .81	.44 .66	.89 .85	.44 .06	.43 .28	.80 .83						2.39 (3.25) 8.08 (4.20)
9. Base 10 T2	.15 .70	.25 .66	.80 .79	.33 .80	.01 .15	.02 .34	.10 .72	.31 .70					0.52 (1.13) 4.92 (4.58)
10. ASPENS comp. T2	.88 .69	.87 .68	.46 .76	.94 .80	.41 .18	.38 .27	.93 .93	.94 .91	.38 .90				8.04 (10.78) 26.99 (22.68)
11. 20 Number line T2	.25 .08	.24 .07	.25 .12	.20 .10	.61 .57	.47 .30	.27 .04	.17 .10	.18 .29	.20 .17			5.22 (2.58) 3.91 (1.83)
12. 100 Number line T2	.44 .28	.47 .28	.01 .35	.43 .34	.61 .42	.72 .70	.48 .36	.40 .32	.06 .38	.42 .37	.55 .22		30.53(12.08) 23.02(11.43)
13. EasyCBM pctile T3	.22 .40	.23 .47	.40 .50	.29 .51	.13 .22	.20 .40	.26 .53	.20 .53	.37 .78	.32 .60	.27 .18	.36 .48	17.67 (16.84) 24.68 (24.57)

Note. Kindergarten statistics are followed by first grade statistics in each cell. Correlations calculated using pairwise deletion. Statistically significant correlations ($p < .05$) are bolded. ASPENS = Assessing Student Proficiency in Early Number Sense; CBM = curriculum-based measurement.

Table 2. Hierarchical Regression Analysis Predicting Fall EasyCBM Percentile With Time 2 ASPENS Composite Scores and NLA 0–100 for First Grade Sample ($N = 39$).

Step	Predictor	R^2	ΔR^2	B	$SE B$	β
1	Intercept	.36	—	7.10	4.41	
	ASPENS composite			0.68	0.13	0.60***
2	Intercept	.49	.13	16.79	11.02	
	ASPENS composite			0.74	0.17	0.56***
	100 number line			-0.62	0.32	-0.25*

Note. CBM = curriculum-based measurement; ASPENS = Assessing Student Proficiency in Early Number Sense; NLA = number line assessment.

* $p < .05$. ** $p < .01$. *** $p < .001$.

were receiving services due to being identified as at-risk in reading. The sample was lower performing in mathematics with average math scores at the 17th (kindergarten) and 25th (first grade) percentile. The composition of the sample, students eligible for summer school services, may have contributed to lower correlations than in previous studies of early numeracy measures and affected the amount of variance accounted for in the regression models. Given the relative complexity of the NLA, results may have differed if a broader set of students had been assessed. Future research should examine the NLA with greater sample sizes, a wider range of student performance, and include data points from across the school year rather than within a limited window during the summer. In addition, the study focused on the value of the NLAs when added to a specific screening battery, ASPENS, future work should include a range of early numeracy screeners. The use of an iPad and the construction of the task required some degree of fine motor skill and thus fine motor skills and other factors such as hand–eye coordination may have been confounding factors affecting student performance. Finally, we were not able to provide sample specific demographics limiting the generalizability of the findings.

The NLA measures used in the current study aligns most closely to the number line task used by Siegler and colleagues, which they hypothesize measures students' knowledge of number magnitude. Using this task, Siegler and colleagues (Booth & Siegler, 2006; Fazio et al., 2014; Laski & Siegler, 2007; Siegler & Opfer, 2003) found that as students develop a greater understanding of magnitude their placement of numbers on a number line transitions from a logarithmic (i.e., estimates exaggerate differences between smaller numbers and compress differences between larger numbers) to linear representation (i.e., differences are neither exaggerated or compressed). This logarithmic to linear shift has been demonstrated across several studies and replicated across grade levels when varying degrees of difficulty are introduced (e.g., Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Geary et al.,

2008; Siegler & Booth, 2004). This shift from logarithmic to linear estimates on the number line task is hypothesized by some researchers to reflect a shift toward a greater understanding of number and magnitude (e.g., Siegler & Opfer, 2003).

Other researchers postulate that the pattern seen in Siegler's number line task does not indicate a shift in students' mental representation of a number line, but instead is an artifact of the task structure. Researchers with this view hypothesize that the upper bound on the number line places a limit on students' estimates of larger target numbers and forces them to shift those estimates down, resulting in the log-linear pattern (e.g., Barth & Paladino, 2011; Cohen & Sarnecka, 2014). To test this theory, Cohen and Sarnecka (2014) examined student performance on an unbounded number line task which included a lower but not an upper bound. They found that the logarithmic to linear shift in student responding was not present. These researchers propose that the unbounded number line task captures students' understanding of number magnitude, while the bounded task used by Siegler and colleagues is first solved through students' use of measurement skills and eventually solved through their engagement in proportional reasoning and use of subtraction and division.

While these viewpoints reflect perspectives of number line task performance within a developmental context, the different task types may have utility as screeners for students at-risk across mathematics achievement levels and implications for research into the utility of using number line assessments as part of screening batteries. Given the depth of work done on bounded number line to date, we used the bounded number line in this study, however future research could tease out how well different number line tasks work with students of varying skill levels and at different grades. For example, if the bounded number line task of Siegler and colleagues taps into more advanced mathematical skills such as subtraction and division, this screener may be better suited to differentiate among students with more advanced mathematics skills and/or for use as a universal screening at later grade levels. In contrast, the unbounded number line task may be more effective at discriminating between at-risk students with lower mathematics skills and/or for use as a universal screener in early elementary. Given the lower performing sample in this study, the bounded number line task may have been too difficult and a better number line screener would be an unbounded "easier" number line task. More closely aligning the measure to the appropriate skill or grade level would help eliminate floor or ceiling effects that might occur due to the difficulty level of the task. Future research could contrast bounded and unbounded tasks at various grades and with varying skill levels to help determine if certain types of number lines work better at different points in time with different populations.

By exploring new tasks and their value in enhancing early numeracy screening systems, the work summarized in this manuscript meets the call to advance the research on screening systems for use in multitier service delivery frameworks (Methe et al., 2011). We would consider it a worthy endeavor to analyze number line tasks in isolation to further the field's understanding of mathematical development. In addition, we also encourage research and researchers to analyze the number line in the context of current screening practice. To that end, research investigating the value added of number line assessments to existing screening batteries should be conducted and findings considered within the framework of cost (e.g., efficiency) to benefit including key decisions related to classification accuracy of screeners (Clarke et al., 2011) and the capability to identify potential nonresponders to interventions (Compton et al., 2012). Advances in this regard will assist the field in furthering our understanding of mathematics development and designing more effective MTSS frameworks and in doing so will, hopefully, help ensure that all children acquire the critical mathematics knowledge necessary for success inside and outside of school.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Ben Clarke, Mari Strand Cary, and Lina Shanley are eligible to receive a portion of royalties from the University of Oregon's distribution and licensing of certain KTEK-based works. Potential conflicts of interest are managed through the University of Oregon's Research Compliance Services.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the *KinderTEK Project*, Grant No. R324A110286, funded by the U.S. Department of Education, Institute of Education Sciences and the *Stepping Up KinderTEK Project*, Grant No. H327S140019, funded by the Office of Special Education Programs, Office of Special Education and Rehabilitative Services. The opinions expressed are those of the authors and do not represent the views of the Institute, the U.S. Department of Education, the Office of Special Education Programs or Office of Special Education and Rehabilitative Services.

References

- Anderson, D., Alonzo, J., & Tindal, G. (2010). *easyCBM® mathematics criterion related validity evidence: Oregon state test* (Technical Report No. 1011). Eugene: Behavioral Research & Training, University of Oregon. Retrieved from http://www.brtprojects.org/wp-content/uploads/2016/05/TechRpt1011_MatheasyCBMOAKS.pdf
- Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. M. (2015). *Evaluation of response to intervention practices for elementary school reading* (NCEE No. 2016-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://files.eric.ed.gov/fulltext/ED560820.pdf>
- Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14*, 125–135. doi:10.1111/j.1467-7687.2010.00962.x
- Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38*, 333–339. doi:10.1177/00222194050380040901
- Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology, 46*, 545–551. doi:10.1037/a0017887
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*, 189–201. doi:10.1037/0012-1649.41.6.189
- Case, R., Okamoto, Y., Griffin, S., McKeough, A., Bleiker, C., Henderson, B., . . . Keating, D. P. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development, 61*, 1–295. doi:10.2307/1166077
- Clarke, B., Doabler, C. T., & Nelson, N. J. (2014). Best practices in mathematics assessment and intervention with elementary students. In P. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (6th ed., Vol. 1, pp. 219–232). Bethesda, MD: National Association of School Psychologists.
- Clarke, B., Doabler, C., Smolkowski, K., *Kurtz Nelson, E., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness, 9*, 607–634. doi: 10.1080/19345747.2015.1116034
- Clarke, B., Doabler, C. T., Strand Cary, M., Kosty, D. B., Baker, S. K., Fien, H., & Smolkowski, K. (2014). Preliminary evaluation of a tier-2 mathematics intervention for first grade students: Utilizing a theory of change to guide formative evaluation activities. *School Psychology Review, 43*, 160–177.
- Clarke, B., Gersten, R. M., Dimino, J., & Rolffhus, E. (2011). *Assessing student proficiency of number sense (ASPENS)*. Longmont, CO: Cambium Learning Group, Sopris Learning.
- Cohen, D. J., & Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology, 50*, 1640–1652. doi:10.1037/a0035901
- Common Core State Standards Initiative. (2010). *Common core standards for mathematics*. Retrieved from <http://www.corestandards.org/the-standards/mathematics>
- Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., . . . Bouton, B. D. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities, 45*, 204–216. doi: 10.1177/0022219412442151
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219–232. doi:10.1177/001440298505200303

- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446. doi:10.1037/0012-1649.43.6.1428
- Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities, 46*, 166–181. doi:10.1177/0022219411410233
- Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology, 123*, 53–72. doi:10.1016/j.jecp.2014.01.013
- Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education, 41*, 121–139.
- Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide* (Practice Guide No. NCEE 2014-4005). Washington, DC: National Center for Education and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/early_math_pg_111313.pdf
- Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*, 493–513. doi:10.1037/0022-0663.97.3.493
- Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities, 45*, 257–269. doi:10.1177/0022219412442167
- Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., 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. doi:10.1177/001440290707300303
- Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33*, 277–299. doi:10.1080/87565640801982361
- Gersten, R. M., & Chard, D. J. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education, 33*, 18–28. doi:10.1177/002246699903300102
- Gersten, R. M., Beckmann, S., Clarke, B., Foegen, A., March, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools* (Practice Guide Report No. NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_math_pg_042109.pdf
- Gersten, R. M., Clarke, B., Jordan, N., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Exceptional Children, 78*, 423–445. Retrieved from <http://cec.metapress.com/content/B75U2072576416T7>
- IBM Corp. (2011). IBM SPSS statistics for Macintosh (Version 20.0). Armonk, NY: Author. Retrieved from <https://www-01.ibm.com/software/analytics/spss/products/statistics/>
- Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences, 20*, 82–88. doi:10.1016/j.lindif.2009.07.004
- Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology, 116*, 45–58. doi:10.1016/j.jecp.2013.02.001
- La Paro, K. M., Hamre, B. K., Locasale-Crouch, J., Pianta, R. C., Bryant, D., Early, D., . . . Burchinal, M. (2009). Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Education and Development, 20*, 657–692. doi:10.1080/10409280802541965
- Laski, E. V. (2013). Preschool and kindergarten portfolio picks: An approach for developing children's metacognition. *YC Young Children, 68*(3), 38–43. Retrieved from <http://www.jstor.org/stable/ycyoungchildren.68.3.38>
- Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 78*, 1723–1743. doi:10.1111/j.1467-8624.2007.01087.x
- Lembke, E. S., McMaster, K. L., & Stecker, P. M. (2010). The prevention science of reading research within a response-to-intervention model. *Psychology in the Schools, 47*, 22–35.
- Methe, S. A., Hojniski, R., Clarke, B., Owens, B. B., Lilley, P. K., Politylo, B. C., . . . Marcotte, A. M. (2011). Innovations and future directions for early numeracy curriculum-based measurement. *Assessment for Effective Intervention, 36*, 200–209. doi: 10.1177/1534508411414154
- Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*, 306–321. doi:10.1177/0022219408331037
- Morgan, P. L., Hillemeier, M. M., Farkas, G., & Maczuga, S. (2014). Racial/ethnic disparities in ADHD diagnosis by kindergarten entry. *Journal of Child Psychology and Psychiatry, 55*, 905–913. doi:10.1111/jcpp.12204
- Muthén, L. K., & Muthén, B. O. (2013). *Mplus user's guide* (Version 7.1). Los Angeles, CA: Author.
- National Assessment of Educational Progress. (2015). *The nation's report card: Mathematics 2015: National assessment of educational progress at grades 4 and 8*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://www.nationsreportcard.gov/reading_math_2015/#?grade=4
- National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel*. Washington, DC: U.S. Department of Education.
- National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: Mathematics Learning Study Committee.

- Salvia, J., & Ysseldyke, J. E. (2004). *Assessment in special and inclusive education* (9th ed.). Boston, MA: Houghton Mifflin.
- Schumacher, R. F., Edmonds, Z. R., & Arden, S. V. (2017). Examining implementation of intensive intervention in mathematics. *Learning Disabilities Research & Practice, 32*, 189–199. doi:10.1111/ldrp.12141
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428–444. doi:10.1111/j.1467-8624.2004.00684.x
- Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237–250. doi:10.1111/1467-9280.02438
- Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*, 273–296. doi:10.1016/j.cogpsych.2011.03.001
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of cronbach's alpha. *Psychometrika, 74*, 107–120. doi:10.1007/s11336-008-9101-0
- Sood, S., & Jitendra, A. K. (2013). An exploratory study of a number sense program to develop kindergarten students' number proficiency. *Journal of Learning Disabilities, 46*, 328–346. doi:10.1177/0022219411422380
- Strand Cary, M., Laski, E. V., Shanley, L., & Clarke, B. (2014). *KinderTEK number line assessment [iPad assessment]*. Eugene, OR: University of Oregon.
- Thompson, B., Diamond, K. E., McWilliam, R., Snyder, P., & Snyder, S. W. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children, 71*, 181–194. doi:10.1177/001440290507100204