A Counting-Focused Instructional Treatment to Improve Number Sense:An Exploratory Classroom-Based Intervention Study

Jessica F. Shumway

Patricia S. Moyer-Packenham

A Counting-Focused Instructional Treatment to Improve Number Sense: An Exploratory Classroom-Based Intervention Study

Jessica F. Shumway[1] & Patricia S. Moyer-Packenham
Utah State University

Abstract: Developing students' number sense is a critical area of research in mathematics education because of the role number sense plays in early mathematics learning. In particular, cognitive psychology research has pinpointed verbal counting as a number sense construct that is critical in later mathematics achievement. This study explored variations in 7- and 8-year-old students' number sense outcomes as they engaged in a counting-focused instructional treatment for differing durations. Sixty students in three elementary classrooms in the United States participated in the counting-focused instructional treatment. A generalized estimating equations (GEE) analysis showed an associated average increase in test scores for students participating in 9 weeks of the counting-focused instructional treatment as compared to students participating in 3 weeks of the instructional treatment. This study provides preliminary results about an instructional practice, based in numerical cognition theory, for elementary mathematics teachers to facilitate opportunities for students to develop their number sense

Keywords: number sense; counting; elementary mathematics; mathematics instructional practices

Introduction

Although number sense has been described and studied since the early twentieth century (e.g., Brownell, 1945; Dantzig, 1954), researchers continue to investigate its constructs and educators remain interested in its role in students' mathematics understanding (Berch, 2005; National Mathematics Advisory Panel, 2008). Current evidence suggests that early number sense is critical to students' later mathematics achievement (Geary, 2011; Jordan, Glutting, & Ramineni, 2010; Locuniak & Jordan, 2008). This evidence has prompted researchers to develop and test interventions to improve young children's early understanding of number (e.g., Clements,

---

[1] jessica.shumway@usu.edu

Sarama, Spitler, Lange, & Wolfe, 2011; Dyson, Jordan, & Glutting, 2013; Ramani & Siegler, 2008; Wilson, Dehaene, Dubois, & Fayol, 2009). While the evidence points to the importance of early intervention for developing students' number sense, there are few studies that examine whole-class instructional treatments focused on specific constructs of number sense (e.g., counting, subitizing, estimation, computation skills). In this study, we examined the verbal counting construct of number sense by exploring variations in 7- and 8-year-old students' number sense outcomes as they engaged in a whole-class counting-focused instructional treatment for differing durations. This included analyzing variations in student outcomes, and the influence of the instructional treatment at the class, subgroup, and individual levels.

 Number sense in mathematics education and cognitive psychology

Researchers make a distinction between the conceptual definition of number sense and its operational definitions (Berch, 2005). This distinction emphasizes the complexity of number sense (conceptual definition) while providing a framework for researchers to utilize in order to assess number sense (operational definitions). Even within these distinctions, researchers further distinguish number sense as either the broad, educational definition or as the more specific, cognitive psychology definition, both of which have conceptual and operational definitions.

In cognitive psychology, the conceptualization of number sense is based on the idea that humans and nonhuman animals are born with an intuitive sense of quantity (Dehaene, 1997). This involves the ability to quickly perceive small amounts (subitize), approximate numerical magnitudes, and comprehend simple number transformations (Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004; Halberda & Feigenson, 2008). This sense of number is nonverbal, nonsymbolic, and an innate, internal cognitive process.

In mathematics education, the conceptualization of number sense is built on a nonverbal, nonsymbolic definition of number sense, but also includes symbolic representations and understandings of number acquired through formal and informal experiences. For example, much of the research on number sense in mathematics education is focused on formal school experiences that promote counting, exact representations of number, quantities tied to symbols, and number system concepts (e.g., Baroody, Eiland, & Thompson, 2009; Dyson et al., 2013; Malofeeva, Day, Saco, Young, & Ciancio, 2004; Ramani & Siegler, 2008).

Assessing number sense

In order to measure children's number sense outcomes, researchers operationalize constructs of number sense. Current research investigating the predictive relationship between early number sense and later mathematics achievement provides a basis for operationalizing number sense constructs. Jordan, Kaplan, Olah, and Locuniak (2006) used a variety of tasks (e.g., counting,

number recognition, nonverbal calculation) for assessing and examining kindergarten students' number sense development over the course of a school year. Their assessment was used in a longitudinal study and later named the Number Sense Brief measure (Jordan et al., 2010). Counting, number relationships, and basic operations emerged as unique constructs of number sense predictive of later success in mathematics learning.

Similar to the Jordan and colleagues' studies, Geary's (2011) predictive longitudinal panel study identified quantitative competencies of first-grade students that predict mathematics achievement and growth of students through fifth grade using a variety of number sense tasks (Number Sets test, violations of counting rules test, number line task, and numerical operations task). His findings supported Jordan and colleagues' results indicating that counting procedures, number knowledge, and basic operations are particularly important in predicting students' mathematics achievement. Geary (2011) used a Number Sets Test and a number line task. Both assessments moved beyond assessing students' number recognition and naming and, in addition, assessed students' fluency in attaching Arabic numerals to small quantities and students' knowledge of the number line. The findings suggest that mapping numerals onto quantities and mapping numbers onto the mathematical number line may be critical to early number skills that impact later mathematics achievement. Hence, the findings from Geary's (2011) research supported and extended Jordan et al.'s (2010) finding that specific early number sense skills correlated with later mathematics achievement, and further operationalized the complex construct of number sense. Counting, quantity discrimination, number combinations, number identification, and estimation were most commonly used in operationalizing number sense and assessing children's number sense across the measures in these studies.

Number sense interventions

Research on the predictive relationship between early number sense and later mathematics achievement establishes the purpose for exploring and improving number sense interventions in elementary classrooms. Three recent studies (Aunio, Hautamaki, & Van Luit, 2005; Dyson et al., 2013; Jordan, Glutting, Dyson, Hassinger-Das, & Irwin, 2012) tested the effects of number sense interventions in the form of a program (i.e., a set of lessons or activities). In Aunio et al.'s (2005) study, 45 preschool students participated in an intervention in small groups of 5-6 children with approximately 60 sessions. Their intervention was based on two established programs (*Let's Think!* and *Maths!*). Dyson et al. (2013) conducted their intervention with 121 kindergarten students in small groups (consisting of 24 lessons). Jordan et al. (2012) also implemented an intervention with 44 kindergarten students in small groups (24 lessons). In all three studies, precursors to counting and counting skills were a key part of the interventions and number sense measures. In comparing the three interventions, number sense was conceptualized as embodying a thinking ability in the Aunio et al. (2005) study, while Dyson et al. (2013) and Jordan et al. (2012) delineated specific skills tied to number sense and operationalized the construct.

Two other recent studies (Ramani & Siegler, 2008; Wilson et al., 2009) tested the effects of number sense interventions in the form of a game that focus on a specific construct of number sense (i.e., a game focused on number line estimation). In Ramani and Siegler's (2008) study, 124 preschool students participated in a board game intervention focused on number line estimation. Their results indicated that the effect of playing a number board game versus a color board game, increased students' proficiency on numerical tasks. Wilson et al.'s research (2009) focused on the concept of *number sense access*, which they defined as the linking of symbolic and nonsymbolic representations of number. They developed an adaptive number sense computer game to improve students' performance on symbolic numerical comparison tasks. They tested the intervention with 53 kindergarten students and assessed their learning with symbolic and nonsymbolic measures for number sense. Results showed improvement on the symbolic numerical comparisons, but no improvement on the nonsymbolic measures. Wilson et al. (2009) concluded that the intervention aided students in their linking of symbolic representations to their representations of quantity. In both studies, one specific construct of number sense was the focus of the intervention.

Overall, some of these recent intervention studies involved multiple components of number sense while others entailed a narrower focus on specific number sense constructs. Although some researchers argue that the constructs of number sense are so intricately interwoven that interventions and assessments cannot and should not isolate and test constructs of number sense (Greeno, 1991), others make the case that isolating key constructs will better identify the constructs that impact number sense instruction, inform future interventions, and provide educators with more information about students' specific difficulties and strengths (Dyson et al., 2013; Ramani & Siegler, 2008; Wilson et al., 2009). Frye et al.'s (2013) review of early numeracy intervention research indicated that numeracy intervention had a positive effect on numeracy achievement, but what was unclear is what aspect of the intervention had the effect or whether specific aspects of numeracy improved. Studies that design and test number sense interventions continue to be needed. A goal of the present research is to contribute to filling this need and examines the verbal counting construct of number sense in the context of a whole-class intervention and its influence on specific aspects of number sense.

Counting as a construct of number sense

Several findings from assessment and intervention studies point to counting as an important component of number sense that impacts students' later mathematics outcomes (Dyson et al., 2013; Jordan et al., 2012; Ramani & Siegler, 2008; Wilson et al., 2009). Counting tasks were commonly used in number sense assessments and as components in the interventions, indicating that counting is a key construct of number sense. Counting facilitates students' understanding of number relationships (Baroody, Eiland, & Thompson, 2009). Wilson et al. (2009) hypothesized

that verbal counting plays an important role in linking nonverbal number knowledge with culturally-based symbols for quantity. As children move beyond the early stages of knowing the verbal count list up to ten and mapping numbers onto larger quantities (LeCorre & Carey, 2007), number sense continues to be important as they progress through the early school grades. These early years of school involve learning a longer count list (beyond 10), becoming more adept with precision of larger quantities, and beginning to link the counting sequences and counting principles to the larger base-ten place-value number system.

While some instructional intervention research in preschool and kindergarten has been conducted by cognitive psychologists and educationalists, research is needed in the early elementary grades when students are using non-symbolic and symbolic understandings of quantity with multi-digit numbers and learning systematic relationships among numbers as the deepen and widen their whole number knowledge. Furthermore, much of the cognitive psychology intervention research was conducted in one-on-one, laboratory, or small group settings. Classroom-based research is needed in order to bridge numerical cognition theory with classroom-based practices. We chose to conduct a classroom-based intervention study in order to facilitate applicable research results while also investigating how and why a counting-focused instructional treatment influences number sense development (Styliandes & Styliandes, 2013).

The present study

The present study investigated the verbal counting construct of number sense and its influence on 7- and 8-year-old students' development of number sense. Based on findings that point to counting as an important component of number sense and predictor of later mathematics outcomes (Geary et al., 2013; Jordan et al., 2012), we hypothesized that verbal counting activities would play an important role in students' number sense outcomes in the short-term. To investigate this hypothesis, we conducted an exploratory classroom-based intervention study of a counting-focused instructional treatment in three classrooms of 7- and 8-year-old students that involved daily verbal counting and discussions about number relationships. Students' number sense outcomes were measured before, during, and after the treatment using assessments of computational fluency, story problems, and number line estimations. The research questions guiding this study were:

What are the variations in number sense outcomes when students engage in a counting-focused instructional treatment for differing amounts of time (3 weeks, 6 weeks, and 9 weeks)?

a) What are the variations in outcomes among three intact classes?
b) What are the variations in outcomes for subgroups of students?
c) What are the variations in outcomes for individual students within each class?

Method

Participants

Sixty 7- and 8-year-old students from three public school classrooms located in one elementary school in the western United States participated in the study. Of the 60 participating students, 52% were male, 48% qualified for free or reduced lunch (indicating low socioeconomic status), and 85% were white. Eight students (13%) had an individualized education plan (IEP) for special education services. Three students (5%) participated in English as a Second Language services.

Participants were assigned to the three classes by school personnel at the beginning of the school year. We did not randomly assign students to instructional treatment groups because our aim was to understand variations among intact classes. Our analysis method controlled for these clusters of students. Instead, we randomly assigned classes to a set length of time for participating in the instructional treatment (Class 1, 9 weeks; Class 2, 6 weeks; Class 3, 3 weeks). Table 1 provides demographic information for the three classes disaggregated by classroom. Overall, Class 1 had the largest percentage of students requiring special education services (27%). There were gender imbalances in Class 2 and 3: 63% of Class 2 was male, and 63% of Class 3 was female. Finally, Class 2 contained the largest percentage of students who qualified for free/reduced lunch, indicating low socioeconomic status (68%). These similarities and variations among participants in the intact classes was an important component of our study because classrooms in the United States have demographic variations within classrooms, schools, districts, and regions.

Table 1

Demographic Characteristics of Participants (N = 60)

Characteristic (or Variable)	Class 1 (*n* = 22)		Class 2 (*n* = 19)		Class 3 (*n* = 19)	
	n	%	*n*	%	*n*	%
Gender						
Male	12	55	12	63	7	37
Female	10	46	7	37	12	63
Socioeconomic (SES) Status						
Low SES	8	36	13	68	8	42
Average/high SES	14	64	6	32	11	58
Race						
White	18	82	17	90	16	84
Black	0	0	1	5	1	5
Hispanic	2	9	0	0	1	5
Asian	2	9	1	5	0	0
Pacific Islander	0	0	0	0	1	5
English as a Second Language (ESL) services						
ESL Services	1	5	2	11	0	0
No ESL Services	21	96	17	90	1	100
Special education services						
IEP	6	27	1	5	1	5
No IEP	16	73	18	95	18	95

Note. Total percentages are not 100 for every characteristic because of rounding.

Procedures, setting, and instructional materials

The study was conducted during the first half of the school year (September to December). Students' number sense was measured using a pretest (prior to instructional treatment), two benchmark tests (during instructional treatment), and a posttest (after instructional treatment). During the instructional treatment phase, we used a pipeline design for staggered treatments (see Figure 1). We did this for two reasons: 1) to provide a comparison among the three classes because a counterfactual was not used; and 2) to explore students' variations in number sense as they participated in the instructional treatment for different amounts of time.

Study Phases	Instructional Treatment		
Pretreatment	Pretest (Classes 1, 2, & 3)		
Week 1	Class 1 begins instructional treatment		
Week 2			
Week 3			
Week 4	Benchmark #1 (Classes 1, 2, &3)		
Week 5	Class 1 continues instructional treatment	Class 2 begins instructional treatment	
Week 6			
Week 7			
Week 8	Benchmark #2 (Classes 1, 2, &3)		
Week 9	Class 1 continues instructional treatment	Class 2 continues instructional treatment	Class 3 begins instructional treatment
Week 10			
Week 11			
Posttreatment	Posttest (Classes 1, 2, & 3)		

Figure 1. Pipeline design for staggered instructional treatment and comparison groups.

Students participated in the counting-focused instructional treatment three days per week during 15-25 minutes of each class's regularly scheduled mathematics block of time. The first author taught all the instructional teaching episodes, while the classroom teachers continued to use their district-adopted curriculum materials for planning and teaching regular mathematics lessons. The researcher conducted the counting-focused instructional treatment in the meeting

area of each classroom while students sat in a circle on the rug. Each of the 27 teaching episodes for the counting-focused instructional treatment followed a fairly standard format using the *Count Around the Circle* number sense routine (Shumway, 2011; Shumway & Kyriopolous, 2013). Count Around the Circle is a routine that involves whole-class participation, with each child saying a number as the class counts through a counting sequence around the circle.

Within each teaching episode, there was a verbal counting component, written counting component (on a number line or grid), and a discussion around number sense concepts (e.g., relations among numbers or estimation). The researcher used a variety of counting sequences, for example, count by tens starting at 57. One student counted on by ten and said "sixty-seven," the next student said "seventy-seven," and so on, until students counted all the way around the circle.

Each teaching episode included some type of symbolic or non-symbolic representation. Most teaching episodes involved the use of a number grid, an open number line, written counting sequences, and/or other visual materials to highlight key ideas and students' strategies and ideas about patterns and relationships among numbers. The researcher used a large tablet for recording daily sequences and discussions during instruction. Additionally, students had opportunities to write counting sequences in their Counting Journals as a way to individually solidify understanding and/or reflect on the counting sequence and discussions for the day.

The researcher also facilitated a classroom discussion (Chapin, O'Conner, & Anderson, 2009) with students about the counting sequences, eliciting talk about number relationships and related ideas such as patterns in numbers, place value, decomposing and composing numbers, estimation, and computation.

Based in a constructivist epistemology, each counting sequence was planned to specifically highlight a topic or big idea that children were anticipated to construct based on the planned verbal counting sequence and questions for facilitating discussion. For example, the counting sequences "count by tens starting at zero" and "count by fives starting at zero" were used to highlight the doubling and halving relationships among numbers in the sequences. Each teaching episode included one to three counting sequences, a plan for how to write the sequences, and questions to facilitate classroom discussion.

Measures

Data were collected using three types of whole-class assessments: a pretest, two benchmark tests, and a posttest. The pretest scores served as baseline data and were administered to all three classes prior to Class 1's instructional treatment. The first benchmark test collected data on how students' learning in Class 1 changed during the first three weeks of the instructional treatment. The first benchmark test also provided comparison data for the other two classes that had not yet received the instructional treatment. The second benchmark test provided information on

students' progress in Classes 1 and 2, while providing comparison data on Class 3. The posttest provided data on students' learning progress in all three classes at the conclusion of the instructional treatments.

Drawing upon Geary's (2011) and Jordan et al.'s (2010) work, number sense was operationalized to include basic operations and number relationships (i.e., number line estimation). Three subtests comprised the assessments in this study: (1) The Assessment of Math Fact Fluency, (2) Story Problem Situations, and (3) Number Line Tasks.

Assessment of Math Fact Fluency

The Assessment of Math Fact Fluency (Fuchs, Hamlett, & Powell, 2003) is a battery of addition and subtraction problems (sums up to 18 and minuends up to 18) that measure computational fluency. Students had one minute on each fluency measure (25 problems each) to complete as many problems as they could with a pencil. The tasks were scored as the number correct out of 100 total problems.

Story problem situations

The story problem situations section included four different cognitively guided instruction problem types (Carpenter et al., 1999; Hiebert et al., 1997). Story problem situations were used to understand how students used their number sense foundations to solve problems (Shumway, Westenskow, & Moyer-Packenham, 2016). Students were presented with four different story problem situations (multiplication, part-part-whole, subtraction, and join-change-unknown). Students' written responses to the problem were scored as correct or incorrect. The overall score was calculated as the percent correct out of 4.

Number Line Tasks

The Number Line Tasks assess students' knowledge of the number line and their estimation abilities, specifically their understanding of where numbers fall in relation to one another. Students were presented with a series of blank number lines from 0 and 100 and a target number. They were asked to locate the position of 5 different target numbers (e.g., 64) on the blank 0 to 100 number line by marking a line where the target number belongs. Geary (2011) explained that children's marks on the number lines may reflect how they represent approximate large numerical magnitudes. His findings suggest that mapping numbers onto the mathematical number line may be critical to early number skills that impact later mathematics achievement. The Number Line Tasks draw upon students' understandings of the links between their nonverbal number knowledge and the symbols used to represent this knowledge. The 5 tasks were scored correct if the students' marks fell within an absolute difference of 5 between the

mark and the correct position of the number. The overall score for this subtest was the percent correct out of 5 trials.

Composite score

The Assessment of Math Fact Fluency, Story Problem Situations, and Number Line Tasks were scored separately then combined into a composite score, with each subtest weighted equally. The composite scores were used as the key outcome and served to assess students' pretest to posttest number sense growth and highlight variations in number sense outcomes among classes, subgroups of students, and individual students.

Explanation of the analysis

Researchers conducted a visual inspection of the data combined with the Shapiro-Wilk test of normality on all three classes' test scores to look at the distribution. Line graphs with mean test scores across time points for each class and each individual student were used to visually show variations in number sense outcomes. These line graphs were also used to answer the research questions from the perspective of individual student test scores across time within each of the three classes.

To answer the research questions, we conducted a Generalized Estimating Equations (GEE) analysis for overall performance on the measures to determine variations in test score outcomes among the three groups (Classes 1, 2, and 3). The study involved multiple observations (pretest, benchmark 1, benchmark 2, and posttest) collected from individual students nested in three different classrooms. This clustered data (by class) with repeated measurements of students necessitated a statistical analysis framework capable of handling data within clusters that are correlated, making GEE the most appropriate method of analysis for this type of clustered data (Hardin, 2005).

While a two-way repeated measures analysis of variance (ANOVA) is typically used in educational research to determine whether there are significant differences between the test score means of three unrelated groups across measurement points, the data in this study violated assumptions for ANOVA, including random assignment of participants to treatment groups. Due to the violations of assumptions for ANOVA, a more sophisticated model was needed to analyze students' test scores within clustered classes across measurement points.

Another analysis option, generalized linear models (GLM) with repeated measures, may provide a more appropriate analysis of this type of data, however, this technique assumes that observations are independent. In this study, each participant's data were not independent at each time point because the observations were close together in time. Furthermore, lack of random assignment could lead to the test scores being impacted by teacher effects and/or students' regular interaction with each other.

An analysis method was needed that could describe changes in groups of students' test scores and explore the associated effect of variables, such as time participating in the instructional treatment, while controlling for non-independent observations. Hence, the GEE analysis, which is based on GLM, is a procedure designed for repeated measures yet controls for a lack of independence and takes into account this possible within-group correlation (Ghisletta & Spini, 2004; Hardin, 2005). While GEEs are not frequently used in mathematics education research, Ghisletta and Spini argued:

> …data naturally organized within hierarchies or from longitudinal and panel studies are very frequent in educational and social sciences. For such data, the application of traditional regression models is not adequate; in particular, the statistical dependence arising from the similarity of observations organized within the same cluster, or stemming from the same participant assessed repeatedly, necessitates analyses that do not assume such dependence to be zero. (p. 431)

Results

Model assumptions

The data at the pretest were approximately normally distributed, indicated by the results of the Shapiro-Wilk Test of Normality (Razali & Wah, 2011): Class 1, $p = 0.63$; Class 2, $p = 0.07$; Class 3, $p = 0.34$. Additionally, a visual analysis of the histograms for each class showed an approximately normal curve, and an analysis of the box plots for each class showed the spread of data were approximately symmetrical.

Variations in test scores across time points among the three classes

We used descriptive statistics and line graphs to analyze overall trends in the data from all four time points (pretest, benchmark 1, benchmark 2, and posttest). Table 2 presents the means and standard deviations for each measurement point by class. The results of the descriptive analysis suggest that the test score means in all three classes followed a similar pattern of improvement at Benchmark 1, followed by a slight decline at Benchmark 2, and concluded with another improvement at Posttest. The line graphs in Figure 2 show this pattern of test scores visually and further accentuates the striking consistency in terms of one class not out-performing another throughout the study. In other words, despite the movement in mean scores across measurement points for each class, the mean scores tended to have the same distance between each other with Class 2 consistently performing with the highest scores and Class 3 with the lowest scores. Class 1's mean scores are almost identical to the total mean across the grade level at each measurement point. The results suggest similar gains (and regressions) for each class throughout the study.
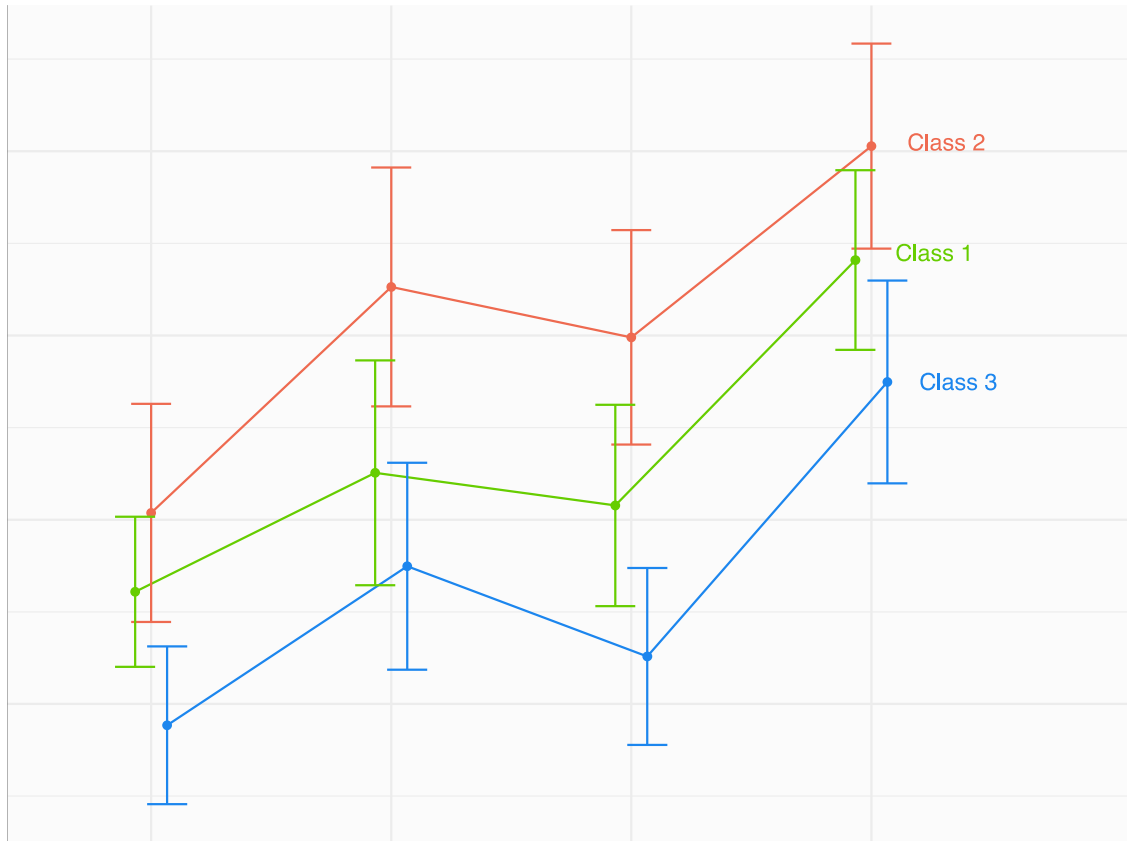
Figure 2. Mean test scores by class at each measurement point.

Table 2

Mean Scores (in Percentages) and Standard Deviations at Each Measurement Point by Class

Class	Pretest		Benchmark 1		Benchmark 2		Posttest	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1 (*n* = 22)	.36	.19	.43	.29	.41	.26	.54	.23
2 (*n* = 19)	.40	.26	.53	.28	.51	.25	.60	.24
3 (*n* = 19)	.29	.19	.37	.24	.33	.21	.47	.24
Total (*n* = 60)	.35	.21	.44	.27	.41	.25	.54	.24

The standard error (SE) bars in Figure 1 overlap at each measurement point suggesting the differences between the means in each class are not statistically significant. This was confirmed from a one-way ANOVA analysis on the pretest scores which indicated that there were no statistically significant differences between group means at pretest ($F(2, 2.60) = 1.42$, $p = .250$).

While the line graphs and descriptive statistics provide some information about the variations in each class's test scores across measurement points, the analysis did not provide results that explained to what degree the classes' variations in performance differed from one another. To answer this research question, we conducted a GEE analysis for overall performance on the measures to determine variations in test scores among the three groups (Classes 1, 2, and 3). The results of the GEE analysis are presented in Table 3. Significant parameters from the GEE analysis included group (i.e., Class 1, 2, or 3), gender, and special education services (IEP). The beta (β) reports a population-averaged parameter representing the averaged effect of a unit change in the predictor for the population, when holding all other variables constant.

Table 3 shows that, when holding all other variables constant, Class 1 had an associated average score of 12.4 percentage points higher than Class 3, which was approaching significance ($β = .12$, $p = .054$). While the GEE results indicate statistical significance at a marginal level, the beta (β) indicates a large effect size in that 12 percentage points higher ($β = .12$) is a whole letter grade higher. Class 2 had an associated average score of 8 percentage points higher than Class 3, which was not statistically significant ($β = .08$, $p = .222$), but approaching a whole letter grade higher. The results from this model suggest that there was an associated increase in test scores when students participated in the counting-focused instructional treatment for longer periods of time (e.g., Class 1 = 9 weeks versus Class 3 = 3 weeks).

Table 3

Generalized Estimating Equations (GEE) Results

Variable	β	SE	95% CI	*p*
Class				
1	.12	.06	-.00, .25	.054*
2	.08	.07	-.05, .22	.222
3	0			
Gender				
Male	.10	.05	.00, .19	.049*
Female	0			
Socioeconomic Status				
Low	.06	.05	-.03, .16	.217
Average/High	0			
Race				
White	-.07	.05	-.17, .03	.179
Black	-.01	.12	-.24, .21	.912
Hispanic	-.20	.13	-.44, .05	.121
Asian	-.16	.11	-.37, .05	.138
Pacific Islander	0			
English as a Second Language				
ESL services	.04	.11	-.17, .26	.7
No ESL services	0			
Special Education Services				
IEP	-.23	.07	-.36, -.10	.001**
No IEP	0			

Dependent Variable: Score.

Model: Class, Gender, SES, Race, ESL, IEP.

* $p < .05$; ** $p < .001$.

Variations in test scores across time points among subgroups

In considering other factors that may influence test scores, the population-averaged parameters showed that students from low socioeconomic (SES) homes scored on average 6.2 percentage points higher than their peers from average/high SES homes when controlling for all other

variables (β = .06, *p* = .217). Though it is not statistically significant, this outcome is atypical of what is generally expected. This outcome could be the result of low power or it could be noise. However, this direction is interesting and possibly with more participants could be important.

The model showed that gender and special education services had significant population-averaged parameters. Table 3 shows that male students had a statistically significant associated average score of 10 percentage points higher than female students when controlling for all other variables (β = .10, *p* = .049). The results also indicated that students with IEPs (i.e., special education services) had an associated average score of 23 percentage points lower than their peers without IEPs, which was statistically significant (β = -.23, *p* = .001). These results suggest that the counting-focused instructional treatment may or may not have been as effective for some female students and some students receiving special education services. Hence, we ran the GEE with subsequent models to investigate interaction effects between class (i.e., Class 1, 2, or 3) and the gender and IEP variables. Building on the base model, we investigated a model with an interaction between class and gender. Similarly, building on the base model, we investigated an interaction between class and IEP.

No significant interactions between gender and class (Class=1*Gender=0, β = -.092, *p* = .461; Class=2*Gender=1, β = .179, *p* = .152) were found, suggesting the effects of the instructional treatment did not depend on gender.

There were significant interactions between class and IEP (Class=1*IEP=0, β = .38, *p* = .002; Class=2*IEP=0, β = .27, *p* = .040). Although Class 1 students with IEPs did better, the result is limited due to single cases of IEPs in Classes 2 and 3 compared to 6 cases in Class 1. Conclusions cannot be drawn from these results; nevertheless, future studies investigating the time that students with IEPs need in these instructional treatments are warranted.

Variations in individual students' test scores within each class

Line graphs of individual student test scores at each measurement point were grouped by class and used to visually show students' variations in number sense development. Figure 3 shows the line graphs for Class 1's individual student test scores. Overall, student learning increased from pretest to posttest. In the path from pretest to posttest, again there was an increase in students' scores at Benchmark 1 followed by a decrease in scores at Benchmark 2. Most of the students' scores followed this pattern, but not all. Of the four students who scored close to 60% or higher on the pretest, two of those students scored lower on the posttests than they did on the benchmark tests. Students who scored between 30% and 55% on the pretest tended to score higher on the posttest than any of the other tests, which was expected.
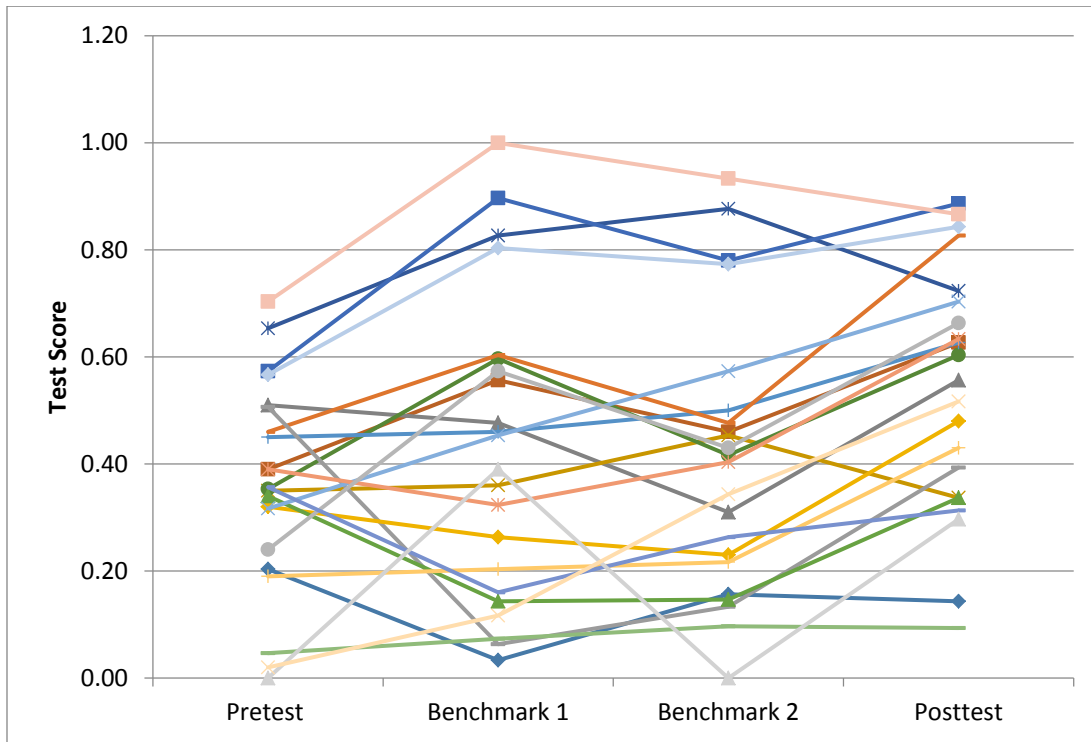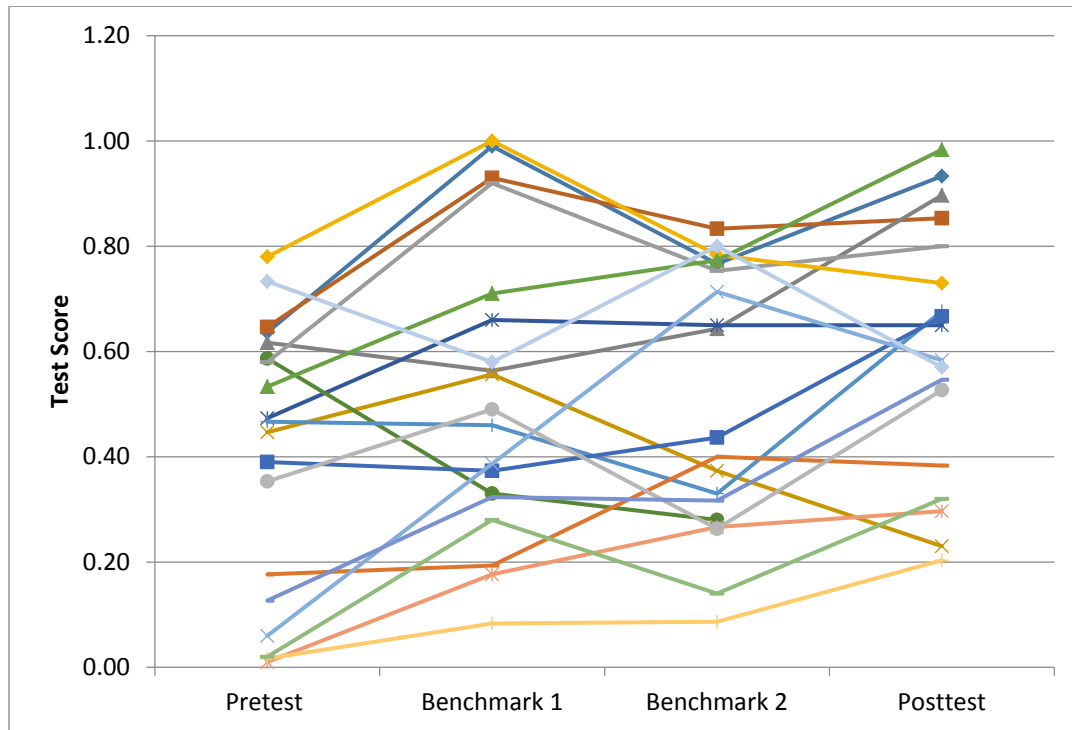
Figure 3. Class 1 individual mean scores.

Figure 4 shows Class 2's individual student test scores across measurement points. Similar to Class 1's trend, three students who scored highest on the pretest scored lower on the posttests than they did on the benchmark tests. Also, similar to Class 1's trend, students who scored in the mid-range of performance on the pretest tended to follow the more expected pattern of higher scores on the posttest, despite a decrease in test scores at Benchmark 2.

Figure 4. Class 2 individual mean scores.

Figure 5 visually highlights that no one in Class 3 scored above 60% on the pretest. Interestingly, students who scored the highest on the pretest in Class 3, followed a similar pattern to the other two classes in that several students scored lower on the posttest than they did on benchmark 1. This pattern remained the same despite starting at a lower score on the pretest. Also notable in Class 3's line graphs is that several students' scores went up by more than 20 percentage points from benchmark 2 to posttest (e.g., from 45% to 70%, from 30% to 65%, from 55% to 83%). That increase occurred after three weeks with the instructional treatment.
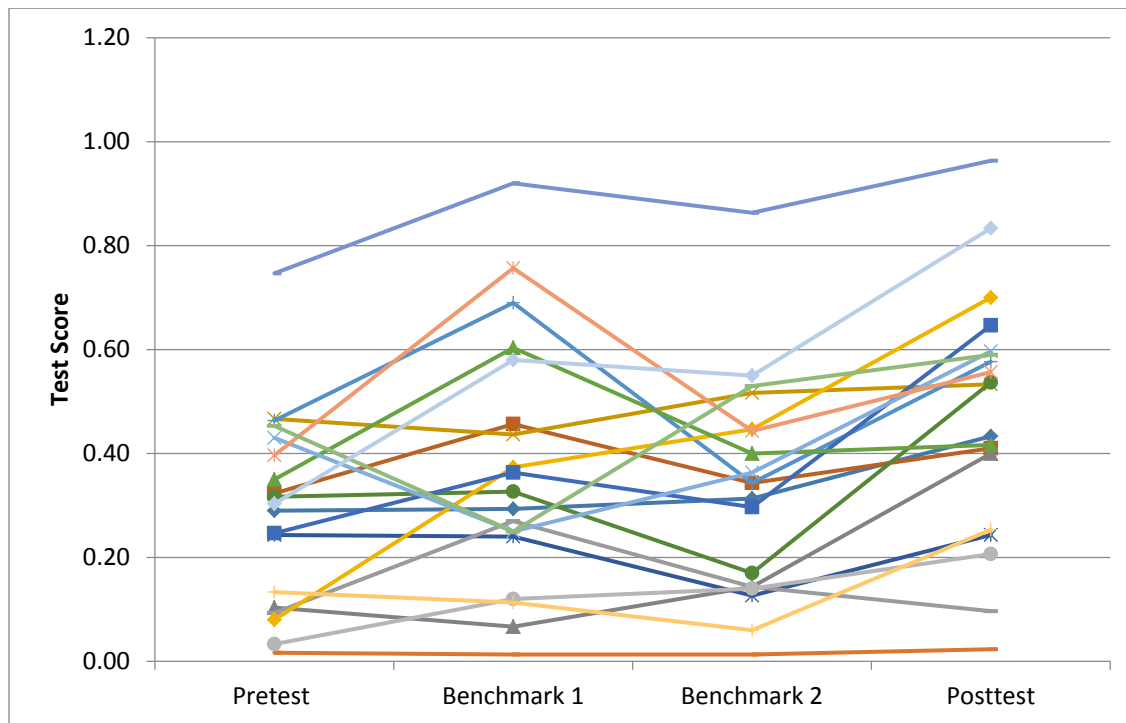
Figure 5. Class 3 individual mean scores.

Discussion

The purpose of this study was to examine the counting construct of number sense by exploring the variations in 7- and 8-year-old students' number sense outcomes as they engaged in a counting-focused instructional treatment for differing amounts of time. The discussion of the results is organized into four sections. In the first section, we describe the variations in number sense development when students engaged in counting-focused instructional treatments for differing amounts of time. The second section identifies the study's limitations. The third and fourth sections present suggestions for future research and conclusions.

Participation in counting-focused instructional treatments and variations in number sense outcomes

This study tested an instructional treatment that was designed to help 7- and 8-year-old students improve their number sense by verbally counting and discussing systematic relationships among numbers in the counting sequences. The results of the GEE analysis showed an associated average increase in test scores when students participated in the counting-focused instructional treatment for longer periods of time. While the GEE analysis cannot attribute the difference in associated average scores of Class 1 and Class 3 solely to more time with the counting-focused

instructional treatment, these results suggest a change in students' outcomes with the counting-focused instructional treatment. Although the study did not control for teacher effects, the evidence suggests that students engaged in 9 weeks of the instructional treatment had differentially better learning outcomes than students engaged in only 3 weeks of the instructional treatment. Future research should incorporate a larger sample, control groups, and controls for teacher effects to determine if the results of the present study hold true beyond the students in this study. Nevertheless, this study provides some initial evidence regarding what dosage (i.e., possibly 9 weeks) of the instructional treatment is needed to be efficacious with 7- and 8-year-old students.

A GEE interaction analysis between group (i.e., Class 1, 2, or 3) and gender was not significant and indicated the instructional treatment worked for males and females alike when holding all variables constant. An interaction analysis between group (i.e., Class 1, 2, or 3) and IEP was significant and indicated that students with IEPs scored lower on average after accounting for the time students were engaged in the instructional treatment. This finding was not surprising, as prior research has shown that students with learning disabilities or learning difficulties tend to score at the 10th to 25th percentile range in most grades, and these patterns of achievement follow them throughout school (e.g., Geary, Hoard, & Bailey, 2012; Jordan, Hanich, & Kaplan, 2003; Murphy, Mazzocco, Hanich, & Early, 2007). This finding was explained by the nature of the interaction between Class (i.e., time participating in the instructional treatment) and IEP.

The interaction analysis line graphs indicated that students with IEPs, on average, did better in Class 1 when they participated in the instructional treatment for 9 weeks versus in Class 2 or 3 when they participated in the instructional treatment for 6 weeks and 3 weeks, respectively. However, conclusions cannot be drawn from these results because of the limited number of students with IEPs and because this did not control for teacher effects. Nevertheless, the longer students with IEPs participated in the counting-focused instructional treatment in this study, the better they performed. This result could have important implications for the inclusion of students with special needs in mainstream classrooms and highlights their opportunities for accessing the content in the whole-class setting.

Finally, because this is an exploratory study, it is worth discussing the results of the GEE analysis in terms of students' socioeconomic status. The results revealed that students from low-socioeconomic backgrounds scored on average 6.2 percentage points higher than their peers from higher socioeconomic backgrounds. Although this finding was not significant, the outcome was atypical of what is generally found in the literature (Clements & Sarama, 2008; Jordan & Levine, 2009). Typically, students from low socioeconomic backgrounds perform worse in mathematics than their peers from higher income families (National Mathematics Advisory Panel, 2008). This result means that the counting-focused instructional treatment could be an

important instructional activity for providing opportunities for students from low socioeconomic backgrounds to develop their number sense. This direction is interesting could be important to investigate in future research for closing achievement gaps in mathematics.

This exploratory study provides important contributions to classroom-based practice and number sense research. For many teachers, it is difficult to orchestrate differentiated, whole-class mathematics instructional activities due to their students' wide-ranging mathematics readiness levels. This study provides some preliminary evidence of a number-sense-focused instructional practice for elementary teachers facilitating whole-class mathematics instruction. The GEE analysis showed that students in this study performed better with 9 weeks of the instructional treatment, students with IEPs had better outcomes with 9 weeks of the instructional treatment, and students from low socioeconomic backgrounds benefited from the instructional treatment. Number sense theory (Baroody & Rosu, 2006; Greeno, 1991; Resnick et al., 1990) indicates that number sense cannot be taught as a lesson or unit of study, rather number sense development is ongoing and requires multiple, connected experiences with number sense ideas. This study provides some initial evidence that engagement in at least 9 weeks of connected number sense experiences, at least 3 days per week, can result in important shifts in learning as students develop their number sense.

Limitations

While the design was methodologically sound and accounted for the complexity of classroom-based intervention research, an intervention study in a classroom setting has limitations. The design did not have a large sample with random assignment and control groups, and therefore, presented limitations to the conclusions and generalizations. The results of the study were interpreted with caution and were not interpreted as causal.

To strengthen the statistical conclusion validity, we used psychometrically sound measures established by previous research (e.g., Fuchs et al., 2003; Geary, 2011; Hiebert et al., 1997). Repeated measurements over time (i.e., pretest, benchmarks, and posttest) compensated for the weaknesses in the statistical design. Since a counterfactual was not used in the study, the pipeline design provided comparison groups.

In addition to threats to statistical conclusion validity, threats to external validity also presented limitations to the instructional treatment study. The study was conducted at one school. Since it was limited to one context, the results are limited in terms of generalizability to other schools or contexts. Location, history, and local teachers, students, politics, and policies would likely affect the outcomes of the instructional treatment in other settings.

Future research

Mathematics education researchers have moved from an initial descriptive research phase in number sense research to prediction (using correlational designs) to improvement (using experimental designs testing various interventions). This study builds on this developing research agenda by designing and testing a counting-focused instructional treatment at the classroom level. This study was implemented at the classroom level in order to increase the likelihood that the research results are applicable to every day, real-life classrooms. However, experimental research on number sense interventions is needed in order to generalize to multiple populations. Experimental research on number sense interventions at the school and district levels could also help to determine if the results of this study were unique to these students or if these learning shifts would be common in the larger population. More specifically, research on how students with IEPs are participating in whole-class practices and what they are accessing would provide educators and researchers with a stronger knowledge base about what types of practices are equitable and effective for all students. One-on-one teaching experiments could lend further insight into the mechanisms for students struggling to develop number sense and the learning paths of students with well-developed number sense.

Conclusion

We proposed that a counting-focused instructional treatment holds promise for refinement in students' numerical precision and understanding of the number system (Le Corre & Carey, 2007). Specifically, 7- and 8-year-old students must use the translation between symbolic and nonsymbolic quantity to begin extending their understanding of the base-ten system and develop fluency with addition and subtraction (CCSSM, 2010). This type of knowledge makes formal mathematics learning more accessible. Studies show that having this knowledge in elementary school predicts better functional mathematical ability in adolescence (Geary et al., 2013). Classroom-based research that bridges understanding between numerical cognition theory and classroom-based practices can help educators to better provide children with opportunities to develop robust number sense. This study provides preliminary results about an instructional practice, based in numerical cognition theory, for elementary mathematics teachers to facilitate opportunities for students to develop their number sense.

References

Aunio, P., Hautamaki, J. & Van Luit, J. E. H. (2005). Mathematical thinking intervention programmes for preschool children with normal and low number sense. *European Journal of Special Needs Education, 20*(2), 131-146.

Baroody, A. J., Eiland, M., & Thompson, B. (2009). Fostering at-risk preschoolers' number sense. *Early Education & Development*, *20*(1), 80-128.

Baroody, A. J., & Rosu, L. (2006, April). *Adaptive expertise with basic addition and subtraction combinations: The number sense view*. Paper presented in symposium at the annual meeting of the American Educational Research Association, San Francisco, CA.

Berch, D. B. (2005). Making sense of number sense: Implication for children with mathematical disabilities. *Journal of Learning Disabilities, 38*(4), 333-339.

Brownell, W. A. (1945). When is arithmetic meaningful? *Journal of Educational Research, 38*(3), 481-498.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: Cognitively guided instruction.* Portsmouth, NH: Heinemann.

Chapin, S. H., O'Connor, C., & Anderson, N. C. (2009). *Classroom discussions: Using math talk to help students learn* (2nd ed.). Sausalito, CA: Math Solutions.

Clements, D.H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*, 443-494.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011.) Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education, 42*(2), 127-166.

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. Retrieved from http://corestandards.org

Dantzig, T. (1954). *Number: The language of science*. New York, NY: MacMillan.

Dehaene, S. (1997). *Number sense: How the mind creates math.* New York: Oxford University Press.

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*(2), 166-181.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307-314.

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide* (NCEE 2014-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website: http://whatworks.ed.gov

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). *Grade 3 math battery.* Nashville, TN: Peabody, Vanderbilt University.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*(6), 1539-1552.

Geary, D. C., Hoard, M. K., & Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities, 45*(4), 291-307.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE 8*(1): e54651.

Ghisletta, P. & Spini D. (2004). An introduction to Generalized Estimating Equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational & Behavioral Statistics, 29*(4), 421-437.

Greeno, J. G. (1991). Number sense as situated knowing in a conceptual domain. *Journal for Research in Mathematics Education*, *22*, 170-218. doi:10.2307/749074

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457-1465.

Hardin, J. W. (2005). Generalized estimating equations. *Encyclopedia of statistics in behavioral science.* Wiley Online Library.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, & Irwin, C. (2012). Building kindergartners' number sense: A randomized controlled study. *Journal of Educational Psychology*. Advance online publication. doi: 10.1037/a0029018.

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning & Individual Differences, 20*, 82-88.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies of children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*, 834-850.

Jordan, N. C., Kaplan, D., Olah, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*(1), 153-175.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850-867.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*, 395-438.

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41*(5), 451-459.

Malofeeva, E., Day, J., Saco, X. Young, L., & Ciancio, D. (2004). Construction and evaluation of a number sense test with Head Start children. *Journal of Educational Psychology, 96*(4), 648-659.

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40*, 458-478.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*, 375-394.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling & Analytics, 2*(1), 21-33.

Shumway, J. F. (2011). *Number sense routines: Building numerical literacy every day in grades K-3*. Portland, ME: Stenhouse.

Shumway, J. F., & Kyriopolous, J. (2013). Mastery multiplied. *Educational Leadership, 71*(4), 73-76.

Shumway, J. F., Westenskow, A., & Moyer-Packenham, P. S. (2016). An exploratory study of a story problem assessment: Understanding children's number sense. *International Journal for Mathematics Teaching and Learning, 17*(3). Retrieved from http://www.cimt.org.uk/ijmtl/index.php/IJMTL/index

Stylianides, A. J., & Stylianides, G. J. (2013). Seeking research-grounded solutions to problems of practice: Classroom-based interventions in mathematics education. *ZDM, 45*(3), 333-342.

Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, & Education, 3*(4), 224-234.

End Note