Learning Disabilities Research & Practice, *22*(1), 36–46

Predicting First-Grade Math Achievement from Developmental Number Sense Trajectories

Nancy C. Jordan
University of Delaware

David Kaplan
University of Wisconsin–Madison

Maria N. Locuniak and Chaitanya Ramineni
University of Delaware

Number sense development was tracked from the beginning of kindergarten through the middle of first grade, over six time points. Children ($n = 277$) were then assessed on general math achievement at the end of first grade. Number sense performance in kindergarten, as well as number sense growth, accounted for 66 percent of the variance in first-grade math achievement. Background characteristics of income status, gender, age, and reading ability did not add explanatory variance over and above growth in number sense. Even at the beginning of kindergarten, number sense was highly correlated with end of first-grade math achievement ($r = 0.70$). Clarifying the observed slope effect, general growth mixture modeling showed that children who started kindergarten with low number sense but made moderate gains by the middle of kindergarten had higher first-grade math achievement than children who started out with similarly low number sense with flat growth. The majority of children in the low/flat growth class were from low-income families. The findings indicate that screening early number sense development is useful for identifying children who will face later math difficulties or disabilities.

Poor achievement in mathematics is a national concern. Although 6–14 percent of the school-age population is estimated to have genuine learning disabilities in math (Barbaresi, Katusic, Colligan, Weaver, & Jacobsen, 2005), many more students are struggling to stay on grade level. Students may have weaknesses in one or more subareas of math because of selective cognitive deficits, inadequate instruction, or a combination of factors (Geary, 2004).

State-mandated math assessments associated with the No Child Left Behind Act of 2001 typically begin in third grade. However, early screening in kindergarten and first grade can identify children in need of educational support or intervention before failure occurs. In virtually every state and school district, children are screened for potential reading difficulties in the primary grades (Gersten & Jordan, 2005). Although reading screens sometimes identify false positives, in other words, children who perform poorly on the screen but go on to achieve normally, the results have been important for identifying those who will need additional support as well as for monitoring progress. Moreover, effective reading screens have led to the development of evidence-based interventions (e.g., interventions targeting phonological awareness in reading; Bus & van IJzendoorn, 1999). In math, on the other hand, research on valid screening for potential math difficulties is in its infancy (Gersten, Jordan, & Flojo, 2005). As a result, children with math difficulties are likely to be underserved in early elementary school.

Most children come to kindergarten with some degree of number sense, although there are wide individual differences (Klibanoff, Levine, Huttenlocher, Vasilyeva, & Hedges, 2006). In fact, even infants are sensitive to changes in quantity (Wynn, 1992). Although the term has been defined differently, most agree that number sense involves abilities related to counting, number patterns, magnitude comparisons, estimating, and number transformation (Berch, 2005). It reflects a child's early experiences, as well as his or her cognitive facility (Dowker, 2005; Lipton & Spelke, 2003). Presumably, number sense lays the foundation for learning formal math concepts and skills in elementary school.

As a part of the Children's Math Project—a prospective longitudinal investigation of the development of math skills in children at risk—Jordan, Kaplan, Olah, and Locuniak (2006) developed a theoretically driven number sense battery for kindergartners. The number sense battery comprised several core areas: *Counting* skills and principles, which looks at knowledge of the count sequence, the ability to enumerate sets, number recognition, and understanding of counting principles (Geary, Hoard, & Hamson, 1999; Gelman & Gallistel, 1978); *Number knowledge*, which involves the ability to compare quantities, such as which of two numbers is larger or smaller (Griffin, 2004); *Nonverbal calculation,* or the ability to perform simple addition and subtraction transformations

Requests for reprints should be sent to Nancy C. Jordan, University of Delaware, School of Education, University of Delaware, Newark, DE 19716. Electronic inquiries may be sent to njordan@udel.edu.

with objects (Hughes, 1986; Huttenlocher, Jordan, & Levine, 1994; Klein & Bisanz, 2000); *Story problems*, which assesses the ability to solve simple word problems where objects were referred to but not presented (e.g., "Jill has two pennies. Jim gives her one more penny. How many pennies does Jim have now?") (Ginsburg & Russell, 1981; Levine, Jordan, & Huttenlocher, 1992); and *Number combinations*, which involves verbally presented calculations with no object referents (e.g., "How much is 2 and 1?"). Number sense was assessed over four time points, from the beginning to end of kindergarten. Pattern recognition and estimation of quantities also were assessed, but the tasks proved to be unreliable.

Jordan and colleagues reported several key results with their kindergarten sample. They found that, compared to their middle-income peers, children from low-income households entered school with a generally low level of number sense. Not surprisingly, caregivers of low-income children reported fewer home experiences with numbers as well as with literacy. All children in the study were exposed to the same math curriculum in kindergarten. Although most kindergartners developed number sense over multiple time points, the income gap did not get smaller with instruction. Analysis of performance in subareas of number sense showed that low-income children had particular difficulties with story problems involving simple arithmetic and that the income gap on story problems widened during the school year. Low-income kindergartners were four times more likely to fall into a low-performing, flat growth category on story problems than middle-income kindergartners. In addition to simple arithmetic, story problems require language comprehension and auditory attention, areas that are particularly sensitive to early experience (NICHD Early Child Care Network, 2005). The analyses further showed small but reliable gender effects in kindergarten number sense, always favoring boys. The effects were present regardless of income level, age, and reading skill.

Remaining questions from Jordan et al.'s (2006) study are the extent to which growth and performance in number sense predict formal math achievement in primary school and whether income and gender effects hold, widen, or attenuate between kindergarten and first grade. Mazzocco and Thompson (2005) analyzed test items on a psychoeducational test battery and found that subsets of items involving number sense (e.g., reading numerals, magnitude judgments, mental addition of one-digit numbers) were accurate in predicting children who would later develop math disabilities. Similarly, Baker et al. (2002) and Clarke and Shinn (2004) found aspects of number sense, such as magnitude comparisons and quantity discrimination, correlate with math achievement. In all of these studies, however, number sense was viewed from a single time point and growth was not assessed.

The present study is a continuation of the Children's Math Project and extends the work of Jordan et al. (2006) in two ways. First, we tracked children's number sense development in first grade with the same population, using portions of the kindergarten battery that were reliable and sufficiently challenging for first graders. This approach allowed us to observe children's growth trajectories longitudinally over six time points, including the transition between kindergarten and first grade. Second, and most important, we measured children's math achievement in first grade. Assessing number sense at multiple time points made it possible to examine rate of growth (slope) as well as level of performance in relation to math achievement. Using the methodology of growth mixture modeling (Muthén, 2004), Jordan and colleagues uncovered unique trajectory paths in kindergarten number sense (e.g., flat vs. steeper growth). We expected prediction accuracy to increase by looking at number sense growth in addition to status. Because income level, gender, age, and reading skill all are associated with math skill (e.g., Jordan et al., 2006; Jordan, Hanich, & Kaplan, 2003a), we considered these variables in our main analyses.

METHOD

Participants

Participants were drawn from a school district in northern Delaware. Children ($n = 414$) were originally recruited for our longitudinal study of children's math in kindergarten. From this group, 277 children remained in the study in first grade when math achievement was assessed. The attrition rate was similar across the six participating schools, primarily due to kindergarten retention or children changing schools after kindergarten. Background characteristics of the participating children at the end of kindergarten (where we set the intercept) and the end of first grade (when children's math achievement was assessed) are presented in Table 1. The demographics are very similar for the two time points. All children were taught math with the *Math Trailblazers* curriculum (Teaching Integrated Mathematics and Science Curriculum, 2004).

Procedure

Children were assessed longitudinally on a number sense core battery in kindergarten (September, November, February, and April) and first grade (September and November). The number sense core battery was the portion of our larger number sense battery (Jordan et al., 2006) that was given to children at all six time points. (In the larger battery, easier tasks were given to children only in kindergarten and harder ones only in first grade.) Children's reading skills were assessed in April of kindergarten and math achievement in April of first grade. Graduate or undergraduate student researchers who were fully trained in test administration and scoring procedures assessed children individually in their schools. The testing took approximately 35 min per child. After each round of testing, the entire research team met to go over scoring and to resolve all questions.

Although the number sense core was given in English, children participating in the English Language Learners program were assessed by a researcher fluent in English and Spanish and allowed to ask that instructions be clarified in Spanish and/or to answer in Spanish. The reading measure, however, was administered to all children in English as prescribed by the school district. Children were tested one school at a time, in approximately the same order for each 1-month testing window.

To ensure accurate data entry, all data were entered twice in the computer. The entries were then checked against each other. If a discrepancy occurred, the error was corrected.

TABLE 1
Demographic Information for Participants at Time 4 ($n = 374$) and Time 8 ($n = 277$)

	Summary	
	Time 4	*Time 8*
Income status		
Low income	123 (33%)	80 (29%)
Middle income	251 (67%)	197 (71%)
Gender		
Male	201 (54%)	151 (55%)
Female	173 (46%)	126 (45%)
Race		
Minority[a]	208 (56%)	142 (51%)
Nonminority	166 (44%)	135 (49%)
English language learners	32 (9%)	27 (10%)
Mean kindergarten start age (*SD*)	67 months (4 months)	67 months (4 months)
Mean reading z score (*SD*)	0.00 (1.00)	
Mean Math Scores (*SD*)		
Calculation standard score		105.54 (15.94)
Applied problems standard score		102.48 (14.80)
Combined raw score		32.54 (6.43)

[a]Minority refers to African American (35%, $n = 130$), Asian (5%, $n = 19$), and Hispanic (16%, $n = 59$) at *Time 4*; and African American (29%, $n = 80$), Asian (5%, $n = 15$), and Hispanic (17%, $n = 47$) at *Time 8*.

Measures

Number Sense Core

Reliability (alpha coefficient) of the number sense core ranged from 0.82 to 0.89 across the six time points. The tasks were always presented to each child in the following order: counting, number knowledge, nonverbal calculation, story problems, and number combinations. The total possible score on the number sense core was 42 points.

Counting

Count sequence, counting principles, and number recognition were assessed with a total possible score of 10 points. For count sequence, children were asked to count to 10 and were given one point if they succeeded in doing so. Children were allowed to restart counting only once but were always allowed to self-correct any number that they were producing. Counting principles were adapted from Geary et al. (1999). For each item, children were shown a set of either five or nine alternating yellow and blue dots. Then a finger puppet told them he was learning to count. The child was asked to indicate whether the puppet counted “OK” or “not OK.” Correct counting involved counting from left to right and counting from right to left. “Pseudo” errors involved counting the yellow dots first and then counting the blue dots or counting the blue dots first and then the yellow dots. For truly incorrect counts, the puppet counted left to right but counted the first dot twice. Children received a score of correct/incorrect for each of eight trials. Finally, children were asked to name a visually presented number. Due to ceiling effects at the end of kindergarten, 3 of the 4 numbers were changed on the battery in first grade. Therefore, only one number, *13*, remained in the core across the six time points.

Number Knowledge

This task was adapted from Griffin (2002) and consisted of eight items. Given a number (e.g., 7), children were asked what number comes after that number and what number comes two numbers after that number. Given two numbers (e.g., 5 and 4), children were asked which number was bigger or which number was smaller. Children also were shown visual arrays of three numbers (e.g., 6, 2, and 5), each placed on the point of an equilateral triangle, and asked to identify which number was closer to the target number (e.g., 5).

Nonverbal Calculation

The nonverbal calculation task was adapted from Levine et al. (1992). The tester and child sat facing one another with 45 × 30 cm mats in front of each of them and a box of 20 chips placed off to the side. The tester also had a box lid with an opening on the side. Three warm-up trials were given in which we engaged the children in a matching task by placing a certain number of chips on the mat in a horizontal line, in view of the child, and told the child how many chips were on the mat. After covering the chips with the box lid, the child was asked to indicate how many chips were hiding under the box lid, either with chips or by saying the number.

After the warm-up, four addition problems and four subtraction problems were presented: $2 + 1$; $4 + 3$; $2 + 4$; $3 + 2$; $3 - 1$; $7 - 3$; $5 - 2$; $6 - 4$. The examiner placed a number of chips on her mat (in a horizontal line) and told the child how many chips were on the mat. The examiner then covered the chips with the box lid. The researcher either added or removed chips (through the side opening) one at a time, and simultaneously told the child how many chips were being added or taken away. For each item, the children were asked to indicate how many chips were left hiding under the box. Addition problems were presented before subtraction problems. Children's errors were corrected on the initial addition and subtraction items. The item was scored as correct if the child displayed the appropriate number of chips and/or gave the appropriate number word.

Story Problems

Children were given four addition and four subtraction story problems. The problems were presented orally, one at a time. The calculations were the same as the ones used in nonverbal calculation. Addition problems were presented before subtraction problems. The addition problems were phrased as follows: "Jill has *m* pennies. Jim gives her *n* more pennies. How many pennies does Jill have now?," while the subtraction problems were phrased: "Mark has *n* cookies. Colleen takes away *n* of his cookies. How many cookies does Mark have now?" Children received a score of correct/incorrect for each story problem.

Number Combinations

We gave four addition and four subtraction combinations using the same calculations that were presented in Nonverbal Calculation and Story Problems. The items were phrased as: "How much is *m* and *n*?" and "How much is *n* take away *m*?" Children were scored as correct/incorrect for each number combination.

Reading

Reading skill was measured with the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) 6th Edition (Good & Kaminski, 2002). The DIBELS assesses letter naming fluency, phoneme segmentation fluency, and nonsense word fluency. The fluency scores are the total number of letters, phonemes, or nonsense words, respectively, identified in 1 minute. Kindergarten reliability was 0.93 for letter naming fluency, 0.88 for phoneme segmentation fluency, and 0.92 for nonsense word fluency. We combined the score for each fluency area to get a total reading score. The three fluency areas are positively and significantly correlated with each other. The total reading score was converted to a *z* score for our analyses.

Math Achievement

Math achievement was measured with the Calculation and Applied Problems subtests of the Woodcock-Johnson III (WJMath; McGrew & Woodcock, 2001). The combined raw scores from these subtests were used as an indicator of math achievement for the analyses. The WJMath is an individually administered, standardized test with national norms. The test is widely used by schools to identify and diagnose learning disabilities (Salvia & Yssledyke, 2004). Reliability at 7 years of age was 0.87 for Calculation and 0.91 for Applied Problems (McGrew & Woodcock, 2001).

Data Analysis Approach

Previous research by Jordan et al. (2006) used conventional growth curve modeling (Raudenbush & Bryk, 2002; Willett & Sayer, 1994) and growth mixture modeling (Muthén, 2004) to study status and change in the development of number sense for children at risk. In the present study, we extended that approach by adding the proximal outcome of first-grade math achievement as measured by the WJMath battery. Adding a proximal outcome extends growth mixture modeling and is referred to as *general growth mixture modeling* (GGMM; Muthén, 2004).

For the purposes of the present study, we adopted the following modeling steps. First, we began with conventional growth curve modeling to establish the general developmental trend in number sense. We centered the status point or intercept at time 4 (end of kindergarten). We next added the proximal outcome of WJMath (first grade) and regressed it on kindergarten status (intercept) and growth rate (slope) to examine how these growth factors predicted WJMath. (For ease of discussion, we refer to the intercepts and slopes for number sense as *growth factors*, unless reference is made to one or the other individually.) In the final step, we added the background measures of gender (male = 1), income status (low income = 1), kindergarten start age (centered at the mean of the sample), and reading to predict WJMath, as well as the growth factors.

Moving to the GGMM analysis, we began by testing whether a two-class model or three-class model for the number sense trajectories provided adequate fit to the data beyond the conventional growth curve model, which is akin to a one-class model. (Prior studies by Jordan et al. (2006) did not reveal the need to examine a four-class model.) We then added the WJMath proximal outcome as in the conventional growth curve case, but for each class. This procedure determined whether there were differences in prediction of WJMath within each class. We examined the demographic makeup of each class through cross-tabulations.

Treatment of Missing Data

A common feature of longitudinal studies is attrition of participants. Attrition, along with simple nonresponse to items on the battery of measures, gives rise to the usual problems of missing data in longitudinal studies. Most ad hoc approaches to dealing with missing data assume that the missing data are "missing-completely-at-random" (MCAR). In essence, the MCAR assumption implies that the process that generated the missing data was completely independent of all variables in the model. That is, the missing data are not generated by either the data that were observed or by the data that are missing. By all accounts, MCAR is a difficult assumption to maintain. At the other end of the spectrum is missing data that are "not-missing-completely-at-random" (NMAR). Here it is assumed that the process that generates the missing data is a function of the variable on which there is missing data. Correct modeling under NMAR requires a specific model for the process that generates the missing data, and developing such a model is complex (Little & Rubin, 2002).

In between MCAR and NMAR lies the assumption of "missing-at-random" (MAR). MAR assumes that missing data can be generated by levels of responses in other variables within the model, but not the variable on which there is missing data. This is a less restrictive assumption than MCAR and does not require a specific model for the process that generates the missing data. Fortunately, it is now possible to incorporate MAR-based approaches to handling missing data within the standard structural equation modeling framework (Arbuckle, 1996). An approach to estimation under MAR was proposed by Muthén and incorporated in Mplus (Muthén & Muthén, 2006), the software program used in this study for the growth mixture modeling.

RESULTS

Correlations between math achievement (as assessed by the WJMath) and the number sense battery and subareas, at the six time points, are presented in Table 2. All of the correlations are positive and significant. The core battery consistently predicts math achievement (e.g., 0.70 at time one, 0.72 at time 6) across the six time points. With the exception of counting, the individual subareas had good predictability, even at the beginning of kindergarten.

Table 3 presents the results of the conventional growth curve models. Model 0 refers to the baseline model with no predictors. Model 1 adds the regression of the math achievement proximal outcome (WJMath) on the growth factors. Model 2 adds the remaining demographic and reading score predictors of WJMath and growth factors. A path diagram of Model 2 is given in Figure 1. The path diagram shows how the repeated measures of number sense relate to the intercept and slope (I and S, respectively, in the diagram) and how the growth factors are predicted by background characteristics. Also, the diagram shows how the background characteristics, as well as the growth factors, predict WJMath.

Referring to Model 0, we see that the growth factors were significantly different from zero, and particularly, the slope was positive, indicating significant and positive growth over the time frame of the study. In addition, the variances of the growth factors were significant, suggesting that there is variance in these parameters that can be explained by the addition of predictors to the model.

TABLE 2
Correlations Between Tasks (at Different Time Points) and WJMath Achievement

	Time 1	*Time 2*	*Time 3*	*Time 4*	*Time 5*	*Time 6*
Number sense	0.70	0.66	0.69	0.73	0.71	0.72
Counting skills	0.36	0.37	0.36	0.35	0.30	0.28
Number knowledge	0.54	0.57	0.52	0.59	0.53	0.54
Nonverbal calculation	0.52	0.40	0.53	0.58	0.50	0.51
Story problems	0.47	0.52	0.54	0.62	0.59	0.62
Number combinations	0.58	0.49	0.58	0.64	0.65	0.68
Reading score				0.51		

Note. All the correlations are significant, $p < 0.01$.

TABLE 3
Parameter Estimates for the Number Sense Growth Curve Models Predicting Performance on WJMath

	Model 0	*Model 1*	*Model 2*
Growth parameters			
Intercept	24.94*	24.96*	25.45*
Slope	0.78*	0.79*	0.79*
Var (intercept)	39.39*	39.17*	21.65*
Var (slope)	0.10*	0.10*	0.09*
Proximal outcome			
WJMath on intercept		0.81*	0.77*
WJMath on slope		2.12*	2.34*
Regression coefficients			
Intercept on male			1.58*
Slope on male			0.06
WJMath on male			0.68
Intercept on low income			−3.42*
Slope on low income			−0.14*
WJMath on low income			0.21
Intercept on K start age			1.07*
Slope on K start age			−0.05
WJMath on K start age			−0.11
Intercept on reading score			3.24*
Slope on reading score			0.02
WJMath on reading score			0.39
BIC	11,989.40	13,616.07	12,807.18
WJMath R^2 statistic		0.66	0.66

Note. Var () stands for the conditional variance of the parameters in parentheses.
$^*p < 0.05$.

Model 1 shows that the growth factors were positively and significantly related to WJMath. Specifically, higher levels of performance in number sense (time 4) were associated with higher levels of performance on WJMath. Moreover, as indicated by the significant slope effect, higher linear rates of growth over the time frame of the study were associated with higher levels of performance on WJMath. Together, the growth factors account for 66 percent of the variance in WJMath.

Model 2 added the background predictors (i.e., gender, income status, kindergarten start age, and reading) of the growth factors and WJMath. Note the growth factors were significant predictors of WJMath, despite adding the background predictors. We found significant differences on the intercept favoring boys, older children, good readers, and middle-income children. The slope effect was significant only for income level, with middle-income children growing at a faster rate. Growth trajectories broken down by gender are shown in Figure 2A and by income level in Figure 2B. The addition of the regression of the WJMath on the background predictors did not appreciably change the percent of variance in math achievement accounted for by the number sense growth factors. Overall, the addition of the background predictors does not add much to the explanatory variance in WJMath over and above number sense.

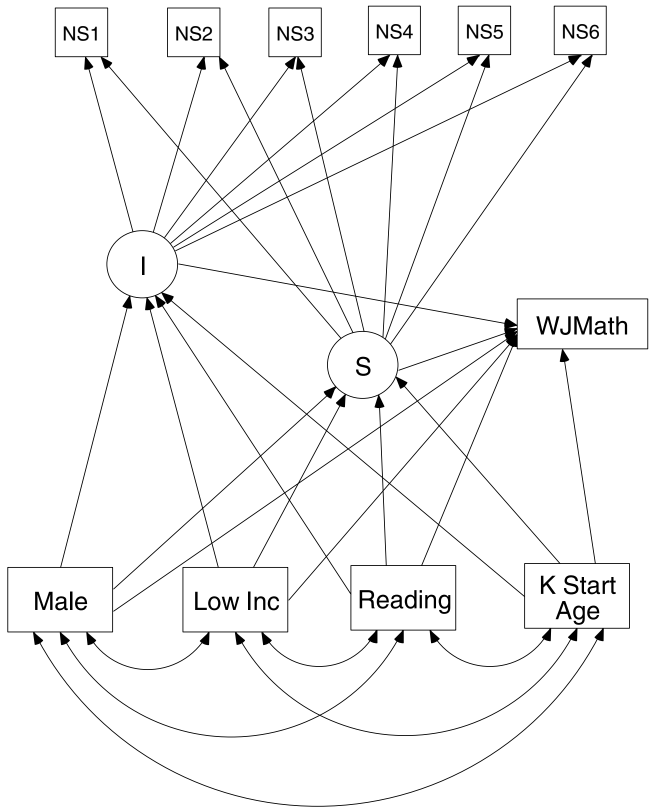

FIGURE 1 Path diagram of conventional growth curve model with background predictors and WJMath as a proximal outcome.

Conventional growth curve modeling assumes that the growth trajectories are a random sample from a single population characterized by a single set of growth parameters. The goal of growth mixture modeling, in contrast, is to relax that assumption allowing for the population to be possibly composed of a finite mixture of subpopulations characterized by their own distinct growth trajectories. Table 4 presents the results of growth mixture modeling applied to the number sense battery from times 1 through 6.

Two-class and three-class models were separately fitted to the data, and on the basis of statistical and substantive considerations, the three-class model was found to fit the data better than the two-class model. Specifically, selecting among models with differing numbers of classes can be determined in three ways. First, one can select the best model by an inspection of the *Bayesian Information Criterion* (BIC). The BIC is a measure that balances the fit of the model with a penalty function for adding parameters to the model. If one attempts to improve the fit of the model by adding a new trajectory class, then the BIC will increase unless the additional trajectory class improves the fit of the model. Thus, we were interested in the smallest BIC among a variety of competing specifications. The model with the lowest BIC value is recommended. Second, one can look at the posterior probability of being assigned to a particular class. In this case, individuals are assigned a probability of being in each of the trajectory classes. The class with the highest probability is the one that the individual is assigned to. A statistical measure that assesses the overall goodness of classification is the *entropy measure* (Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993). Values of entropy approach 1.0 when there is distinct separation of the trajectory classes allowing for clear-cut classification. Finally, one can examine the utility of the number of classes in terms of substantive considerations. For example, prior knowledge of the incidence of problematic math development in the population could be used to determine if certain classes contain reasonable numbers of children. These three approaches were used in this study.

Model 0 of Table 4 is the model without the proximal outcome. We labeled the three classes as *low/flat*, *middle/steep*, and *high/flat*. The labels characterize performance level at the end of kindergarten (where we set the intercept) and the rate of growth (slope). As expected, the classes are similar to the ones found in Jordan et al. (2006). Figure 3 shows the sample and estimated trajectories for the three-class model. Model 1 adds the regression of WJMath on the growth factors. Several findings are worth noting. First, we found that, for each class, better performance in number sense was associated with better performance on the WJMath. However, the rate of growth in number sense was associated with better performance on the WJMath only for the low/flat class. Note that the growth parameters in the low/flat class also explained more of the variance in the WJMath than either the middle/steep class or the high/flat class. Note also that the addition of the WJMath into the model as a proximal outcome did not seriously change the class proportions or the measure of classification adequacy as measured by the entropy measure. This implies that the original three-class model without WJMath remains robust to incorporation of additional information regarding math performance.

Table 5 presents descriptive statistics of children within each of the three trajectory classes. The data corroborate the utility of the three-class solution. Several interesting findings emerged. First, the gender breakdown was roughly even in the low/flat class, whereas boys were the majority among children in the middle/steep and high/flat classes. Not surprisingly, children in the middle/steep and high/flat classes were predominantly middle income whereas the majority of children in the low/flat class were low income. When the data were broken down by income and gender, we found that the distribution of gender by income status for the low/flat class was roughly uniform. By comparison, we found the defining feature of the middle/steep and high/flat classes was income status rather than gender. The average reading score for children in the low/flat class was approximately two-thirds of a standard deviation below the mean, while the average for the high/flat class was about three-quarters of a standard deviation above the mean. The middle/steep class was almost exactly at the mean. Finally, the average WJMath scores for each group showed poorest mean performance in the low/flat group, intermediate performance in the middle/steep group, and highest performance in the high/flat group.

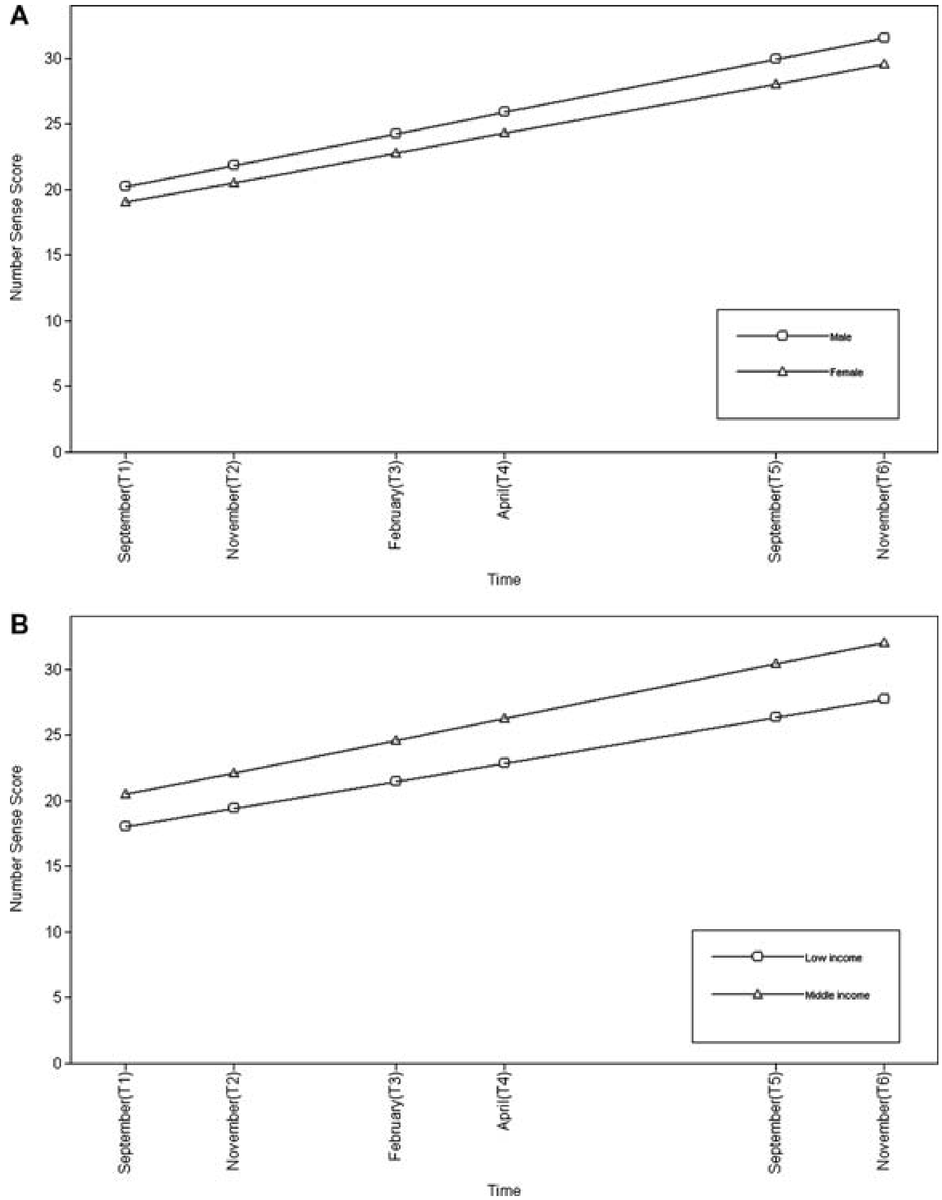

FIGURE 2 (A). Fitted growth trajectories for number sense by gender controlling for income status, kindergarten start age, and reading skill. (B). Fitted growth trajectories for number sense by income status controlling for gender, kindergarten start age, and reading skill.

DISCUSSION

Early number sense is a reliable and powerful predictor of math achievement at the end of first grade. Number sense performance in kindergarten, as well as number sense growth from the start of kindergarten through the middle of first grade, accounted for 66 percent of the variance in first-grade math achievement. Background characteristics of income status, gender, age, and reading ability did not add explanatory variance over and above number sense. Clarifying the observed slope effect, our growth class data (see Figure 3) showed that children who started kindergarten with low number sense but made moderate gains by the middle of kindergarten had higher first-grade math achievement than children who started out with similarly low number sense but remained low. It is somewhat surprising that children, overall, did not make appreciably greater gains in number sense in first grade—with the start of formal arithmetic instruction—than they did in kindergarten. The finding contrasts with those in the reading literature where children made more literacy gains in first grade, when more formal instruction was provided, than they did in kindergarten (McCoach, O'Connell, Reis, & Levitt, 2006). It may be the case that reading receives more emphasis in first grade than math, especially during the first part of the year.

For the purpose of this study, we defined number sense operationally as counting skill, number knowledge, and the ability to transform sets through addition and subtraction.

TABLE 4
Parameter Estimates for the Number Sense Growth Mixture Models Predicting Performance on WJMath

	Model 0			*Model 1*		
	Class #1 Low/Flat	*Class #2 Middle/Steep*	*Class #3 High/Flat*	*Class #1 Low/Flat*	*Class #2 Middle/Steep*	*Class #3 High/Flat*
Growth parameters						
Intercept	18.51*	25.33*	33.51*	18.34*	25.07*	33.44*
Slope	0.51*	1.12*	0.64*	0.48*	1.11*	0.65*
Var (Intercept)	6.17*	6.17*	6.17*	6.31*	6.31*	6.31*
Var (Slope)	0.03*	0.03*	0.03*	0.03*	0.03*	0.03*
Proximal outcome						
WJMath on intercept				1.47*	1.30*	1.06*
WJMath on slope				14.07*	4.11	3.78
BIC		11875.95			13526.64	
Entropy		0.74			0.75	
WJMath R^2 statistic				0.61	0.49	0.39
Final class proportions	0.36	0.39	0.25	0.33	0.42	0.25

Note. Var () stands for the conditional variance of the parameters in parentheses.
$^*p < 0.05$.

Addition and subtraction abilities were assessed in three contexts: nonverbal problems (with physical referents), story problems, and number combinations. It could be argued that story problems and especially number combinations involve conventional arithmetic rather than elementary number sense. In fact, Jordan et al. (2006) found a two-dimensional structure for the number sense battery with number combinations and story problems (conventional verbal arithmetic) clustering on one factor and counting, number knowledge, and nonverbal calculation (more basic number skills) on another. Nonetheless, the strong and significant correlations between performance on number combinations and story problems—even at the beginning of kindergarten—and math achievement suggest that early competencies on these problems tap primary facilities that are fundamental to learning conventional math. Consider what a child must accomplish when solving a verbally presented number combination, before a fact is committed to memory. The child must form a mental representation of a quantity (or represent it with his or her fingers) and then transform the representation through adding or subtracting a specific amount. The child may use visualization or counting or both procedures. This contrasts with the

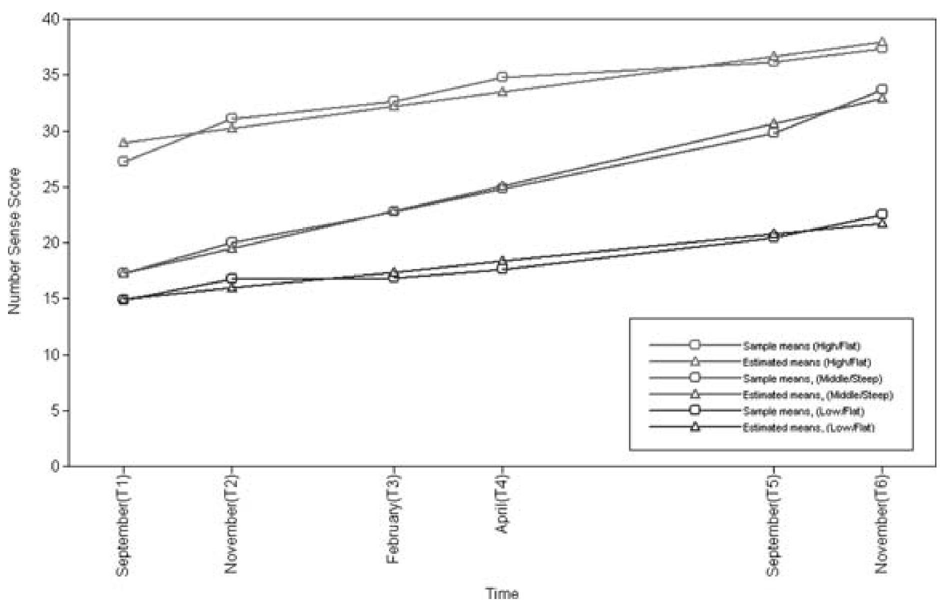

FIGURE 3 Observed and fitted growth trajectories for number sense controlling for gender, income status, kindergarten start age, and reading skill by growth mixture class.

TABLE 5
Number of Children Assigned to Each Class by Gender and Income; and the Mean K Start Age, Reading Scores, and WJMath Scores for Each Class ($n = 414$)

	Class #1 Low/Flat (n = 139)	*Class #2 Middle/Steep (n = 173)*	*Class #3 High/Flat (n = 102)*
Class proportions	0.33	0.42	0.25
Gender			
Female	73 (52%)	77 (44%)	42 (41%)
Male	66 (48%)	96 (56%)	60 (59%)
Income status			
Middle income	63 (45%)	121 (70%)	89 (87%)
Low income	76 (55%)	52 (30%)	13 (13%)
Gender × Income			
Male-low income	34 (25%)	27 (16%)	8 (8%)
Male-middle income	32 (23%)	69 (40%)	52 (51%)
Female-low income	42 (30%)	25 (14%)	5 (5%)
Female-middle income	31 (22%)	52 (30%)	37 (36%)
Mean K start age (*SD*)	66 months (4)	67 months (4)	68 months (4)
Mean reading score (*SD*)	0.63 (0.79)	0.03 (0.84)	0.75 (0.95)
Mean WJMath raw score (*SD*)	27.0 (5.38)	32.55 (4.73)	38.26 (4.19)

Note. Percentages inside parentheses refer to within task model classification. Other values in parentheses indicate standard deviations.

nonverbal calculation task, where the initial amount and the numerical operations are represented physically. To be sure, general cognitive competencies related to language understanding, auditory and visual attention, and working memory support early skill with numbers (Aunola, Leskinen, Lerkanen, & Numi, 2004; Fuchs et al., 2005; Geary, 2004; Klein et al., 2000). However, we argue that the early facility with number combinations depends on primary number sense related to representing, comparing, and manipulating quantities. Such knowledge, along with the ability to infer calculation principles, facilitates fact mastery and reduces the burden on rote memory (Baroody, 1985; Jordan, Hanich, & Kaplan, 2003b). As children associate correct answers for particular number combinations, through counting and/or mental manipulation, they begin to commit combinations to memory and acquire fluent fact retrieval (Siegler & Jenkins, 1987). Of course, whether early facility with number combinations develops intuitively or through particular early experiences or training is an open question.

An issue is the overlap between what children were required to do on the WJMath achievement test and our number sense battery. Clearly, the math achievement test and our battery tapped similar skills. However, the former did so in a pencil-and-paper context that is more appropriate for children after formal instruction is initiated. Not surprisingly, early development of arithmetic, both on informal (e.g., nonverbal calculation) and more school-like (e.g., number combinations) tasks, predicted conventional math achievement. In reading, it has been shown that the early literacy measures that are most directly related to reading (e.g., letter/sound recognition) are the most useful for predicting later reading achievement and providing intervention (Roswell & Chall, 1994). Likewise, arithmetic-like measures at the beginning of kindergarten and associated number sense are predictive of conventional math and should be useful for planning instruction.

Risk Factors

Our diverse sample allowed us to examine the impact of risk factors on children's growth and skill level in number sense. Our low-income participants, defined by participation in the free or reduced lunch program in school, performed worse in number sense relative to their middle-income peers and demonstrated less growth over the study period, while controlling for age, gender, and reading skill. Although both income groups made gains, the trajectory for the low-income group tended to flatten over time. In fact, over 50 percent of the low-income children fell into our low/flat trajectory group versus 12 percent in the high/flat group and 29 percent in the middle/steep group. Basically, many—but not all—low-income children came to kindergarten with weak number sense and gained very little throughout kindergarten and first grade. As reported earlier (Jordan et al., 2006), low-income children start school with a disadvantage in terms of number experiences and thus may be less available for learning math in school. All of the children in the present study were taught with an inquiry-based curriculum, one that does not emphasize direct instruction in number skills. This approach may be less optimal for children who come to school with impoverished number skills and may account for some of the disparities between low and middle-income children. It seems reasonable to suggest that low-income children might benefit from explicit methods targeting number sense.

Over and above income level, we find that gender effects in number sense, favoring boys, extend into first grade. The effects, however, are very small and may have few consequences in practical terms. Like boys, most girls fell in the middle/steep group, followed by the high/flat, and then the low/flat group. The gender patterns within income groups were essentially the same. With respect to reading and age of kindergarten entry, we found that the best readers and the oldest children tended to fall in the high achieving number sense group. In other words, children who come to school with strong number sense tend to be better readers and older. This is not to suggest that teaching children to read will directly influence number sense and math achievement. Rather, early literacy and number sense may be associated with early experiences and have common cognitive antecedents (Aunola et al., 2004). There may be some value in delaying kindergarten entry for children who have borderline, younger ages. However, at this point, it is not clear whether the observed age effects dissipate as children grow older.

In sum, the results of the present study, along with a growing body of research with converging data (Gersten et al., 2005), suggest that screening number sense in kindergarten is useful for identifying children who will face learning difficulties or disabilities in math. To reduce the risk of identifying false positives, children might be screened several times from early to mid kindergarten. Our data indicate that the middle of kindergarten might be a useful time to start providing interventions. In the more established area of reading, early and explicit instruction in phonological awareness leads to achievement gains (e.g., Bus et al., 1999). A plausible next question is whether explicit instruction in number sense can help children compensate for their early shortcomings and whether such instruction can increase math achievement.

ACKNOWLEDGMENT

This study was supported by a grant from the National Institute of Child Health and Human Development (R01 HD36672). We wish to thank the children and teachers who participated in this project.

REFERENCES

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Erlbaum.

Aunola, K., Leskinen, E., Lerkkanen, M. K., & Nurmi, J. E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96*(4), 699–713.

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

Barbaresi, M. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., & Jacobsen, S. J. (2005). Math learning disorder: Incidence in a population-based birth cohort, 1976-1982, Rochester, Minn. *Ambulatory Pediatrics, 5,* 281–289.

Baroody, A. J. (1985). Mastery of basic number combinations: Internalization of relationships or facts? *Journal for Research in Mathematics Education, 16*(2), 83–98.

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

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91,* 403–414.

Clarke, B., & Shinn, M. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*(22), 234–248.

Dowker, A. (2005). *Individual differences in arithmetic: Implications for psychology, neuroscience, and education*. New York: Psychology Press.

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*(3), 493–513.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37,* 4–15.

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology, 74,* 213–239.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38,* 293–304.

Gersten, R., & Jordan, N. C. (2005). Early screening and intervention in mathematics difficulties: The need for action introduced to the special series. *Journal of Learning Disabilities, 38,* 291–292.

Ginsburg, H. P., & Russell, R. L. (1981). Social class and racial influences on early mathematical thinking. Monographs of the Society for Research in Child Development, 46 (6, Serial No. 193).

Good, R. H., & Kaminski, R. A. (2002). *Dynamic indicators of basic early literacy skills (DIBELS), 6th ed*. Eugene, OR: Institute for the Development of Educational Achievement.

Griffin, S. (2002). The development of math competence in the preschool and early school years: Cognitive foundations and instructional strategies. *Mathematical Cognition*, Chapter 1, 1–32.

Griffin, S. (2004). Building number sense with number worlds: A mathematics program for young children. *Early Childhood Research Quarterly, 19,* 173–180.

Hughes, M. (1986). *Children and number difficulties in learning mathematics*. Oxford, UK: Blackwell.

Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General, 123,* 284–296.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003a). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology, 85,* 103–119.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003b). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*(3), 834–850.

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

Klein, J. S., & Bisanz, J. (2000). Preschoolers doing arithmetic: The concepts are willing but the working memory is weak. *Canadian Journal of Experimental Psychology, 54*(2), 105–115.

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher "Math Talk". *Developmental Psychology, 42*(1), 59–69.

Levine, S. C., Jordan, N. C., & Huttenlocher, J. (1992). Development of calculation abilities in young children. *Journal of Experimental Child Psychology, 53,* 72–103.

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science, 14*(5), 396–401.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

Mazzocco, M. M., & Thomspon, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research & Practice, 20*(3), 142–155.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth

during the first 2 years of school. *Journal of Educational Psychology, 98*(1), 14–28.
McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.
Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 345–368). Thousand Oaks, CA: Sage.
Muthén, B., & Muthén, L. (2006). *Mplus user's guide*. Los Angeles: Muthén & Muthén.
NICHD Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology, 41*(2), 428–442.
Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). A empirical pooling approach for estimating marketing mix elasticities using PIMS data. *Marketing Science, 12,* 103–124.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousands Oaks, CA: Sage.
Roswell, F. G., & Chall, J. S. (1994). *Creating successful readers: A practical guide to testing and teaching at all levels*. Chicago: Riverside.
Salvia, J., & Yesseldyke, J. E. (2004). *Assessment in special and inclusive education*. New York: Houghton Miflin.
Siegler, R. S., & Jenkins, E. (1987). *How children discover new strategies*. Hillside, NJ: Erlbaum.
Teaching Integrated Mathematics and Science (TIMS) Curriculum. (2004). *Math trailblazers* (2nd ed.). Chicago: Kendall/Hunt.
Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116,* 363–381.
Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 27,* 749–750.

About the Authors

Nancy C. Jordan is Professor of Education at the University of Delaware. She earned her doctoral degree from Harvard and completed a post doctoral fellowship at the University of Chicago. With support from the National Institutes of Health, she has been studying children with mathematics difficulties for the last 10 years.

David Kaplan is Professor of Quantitative Methods in the Department of Educational Psychology at the University of Wisconsin–Madison. He received his Ph.D. in education from UCLA in 1987. His area of expertise is educational statistics.

Maria N. Locuniak is a doctoral student in school psychology at the University of Delaware.

Chaitanya Ramineni is a doctoral student in educational measurement and statistics at the University of Delaware.