Number Sense Growth in Kindergarten: A Longitudinal Investigation of Children at Risk for Mathematics Difficulties

Nancy C. Jordan, David Kaplan, Leslie Nabors Oláh, and Maria N. Locuniak
University of Delaware

Number sense development of 411 middle- and low-income kindergartners (mean age 5.8 years) was examined over 4 time points while controlling for gender, age, and reading skill. Although low-income children performed significantly worse than middle-income children at the end of kindergarten on all tasks, both groups progressed at about the same rate. An exception was story problems, on which the low-income group achieved at a slower rate; both income groups made comparable progress when the same problems were presented nonverbally with visual referents. Holding other predictors constant, there were small but reliable gender effects favoring boys on overall number sense performance as well as on nonverbal calculation. Using growth mixture modeling, 3 classes of growth trajectories in number sense emerged.

Mathematics difficulties are widespread in the United States as well as in other industrialized nations. The consequences of such difficulties are serious and can be felt into adulthood (Dougherty, 2003; Murnane, Willett, & Levy, 1995). Low math achievement is especially pronounced in students from low-income households (National Assessment of Educational Progress, 2004). Children with weaknesses in basic arithmetic may not develop the conceptual structures required to support the learning of advanced mathematics. Although competence in high-level math serves as a gateway to a myriad careers in science and technology (Geary, 1994), many students never reach this stage. Some children gradually learn to avoid all things involving math and even develop math anxieties or phobias (Ashcraft, 2002; Ginsburg, 1997).

In recent years, there has been growing interest in children with mathematics difficulties, in part because of the relatively large body of research on children's mathematical cognition more generally (Geary, 2000; Jordan, Blanteno, & Uberti, 2003). In general, math difficulties have been defined as below-average performance on a standardized achievement test (Hanich, Jordan, Kaplan, & Dick, 2001). Research on children with mathematics difficulties has been influenced by the wealth of studies in phonetic-based reading difficulties, studies that have led di-

rectly to the development of evidence-based methods for determining who is going to need support and how to provide help (e.g., Foorman, 2003; Torgeson, 2002). As in reading, longitudinal research is critical for understanding how math difficulties develop and change over time (Jordan, Kaplan & Hanich, 2002). For example, Geary and colleagues have shown that some mathematics difficulties are relatively fluid in elementary school whereas others are highly persistent and are characterized by basic cognitive deficits (Geary, 1990; Geary, Brown, & Samaranayake, 1991; Geary, Hamson, & Hoard, 2000). Jordan and colleagues found that although children with mathematics difficulties who are good readers can use their verbal strengths to compensate for some deficiencies in math, they have persistent problems mastering number combinations, such as $3 + 8 = 11$ or $12 - 9 = 3$ (Hanich et al., 2001; Jordan et al., 2002; Jordan, Hanich, & Kaplan, 2003a, 2003b). Difficulties with number combinations reflect basic weaknesses in number sense, such as the ability to grasp counting principles or to manipulate quantities mentally (Geary, Bow-Thomas, & Yao, 1992; Jordan et al., 2003a). There may also be underlying problems in working memory, spatial representation, and attention (Geary, 1994; Mix, Huttenlocher, & Levine, 2002). Longitudinal research on subgroups of mathematics difficulties is beginning to bear fruit for the development of instructional methods. For example, Fuchs, Fuchs, and Prentice (2004) demonstrate that children with mathematics difficulties who are good readers respond differently to

Correspondence concerning this article should be addressed to Nancy C. Jordan, School of Education, University of Delaware, Newark, DE. Electronic mail may be sent to njordan@udel.edu.

instruction than children with mathematics difficulties who are also poor readers.

This investigation was concerned with the emergence of mathematics difficulties. If children's learning needs can be identified early on, we may be able to design interventions that prevent failure in math. The importance of different components of "number sense" to mathematics achievement is not well understood, although the aforementioned research on mathematics difficulties in elementary school is suggestive. Although no two researchers define number sense in exactly the same way (Gersten, Jordan, & Flojo, 2005), most agree that the ability to subitize small quantities, to discern number patterns, to compare numerical magnitudes and estimate quantities, to count, and to perform simple number transformations are key elements of number sense in young children (Berch, 2005; Case, 1998). Most children develop fundamental number sense before they receive formal instruction in elementary school, although there is significant variation by social class and cognitive ability (Ginsburg & Golbeck, 2004; Ginsburg & Russell, 1981; Huttenlocher, Jordan, & Levine, 1994; Jordan, Huttenlocher, & Levine, 1994). Even infants appear to be sensitive to small numbers and number transformations (e.g., Wynn, 1992). Preschool children learn basic counting principles (e.g., Fuson, 1988; Wynn, 1990) and can perform addition and subtraction calculations (e.g., Levine, Jordan, & Huttenlocher, 1992; Mix, Levine, & Huttenlocher, 1999). These foundational aspects of number sense are important to the "higher order" mathematical thinking (e.g., fluency and flexibility with operations and procedures) that results from formal education (Berch, 2005).

In this study, we examined the development of number sense in kindergartners. Our battery, although not exhaustive, assessed skills that have been validated by research and are relevant to the math curriculum in primary school (Griffin & Case, 1997) rather than basic cognitive abilities (e.g., general working memory). The areas included counting, number knowledge, number transformation, estimation, and number patterns and are summarized in Table 1.

Components of Number Sense

Counting

According to Baroody (1987), "Counting puts abstract number and simple arithmetic within the reach of the child" (p. 33). Development of counting is a critical pathway to learning about numbers and

Table 1

Key Elements of Number Sense in Young Children

Area	Components
Counting	Grasping one to one correspondence
	Knowing stable order and cardinality principles
	Knowing the count sequence
Number knowledge	Discriminating and coordinating quantities
	Making numerical magnitude comparisons
Number transformation	Transforming sets through addition and subtraction
	Calculating in verbal and nonverbal contexts
	Calculating with and without referents (physical or verbal)
Estimation	Approximating or estimating set sizes
	Using reference points
Number patterns	Copying number patterns
	Extending number patterns
	Discerning numerical relationships

counting weaknesses have been linked to mathematics difficulties (Geary, 2003). Most children develop knowledge of three important "how to count" principles before they enter kindergarten (Gelman & Gallistel, 1978), including the one–one principle, the stable-order principle, and the cardinality principle. Typically, children learn the count sequence by rote and then discover counting principles through informal experiences with numbers and counting (Briars & Siegler, 1984). As children progress through kindergarten and early elementary school, they gradually acquire more advanced counting abilities. They learn to count backward, to count by twos, and to enumerate object sets greater than 10. They also learn the words for decades and the rules for combining number words (e.g., combining 30 with 3 to make the larger number 33) (Ginsburg, 1989). Counting skills are fundamental to learning base-10 concepts. Early difficulties in counting portend later difficulties with arithmetic operations (Geary, Hoard, & Hamson, 1999).

Number Knowledge

Children as young as 4 years of age recognize and describe global differences in quantities (Case & Griffin, 1990; Griffin, 2002, 2004). For example, they can tell which of two stacks of chips has more or fewer. Younger children rely on visual perception

rather than on counting to make this judgment (Xu & Spelke, 2000). By 6 years of age, however, children integrate their global quantity and counting schemas into a mental number line (Siegler & Booth, 2004). This superordinate structure, referred to as a "central structure for whole numbers," allows children to make better sense of their quantitative worlds (Griffin, 2002). Children gradually learn that numbers later in the counting sequence have larger quantities than earlier numbers. They come to see that numbers themselves have magnitudes, such that 8 is bigger than 5 or that 6 is smaller than 9. Children use these skills in a wide range of contexts and eventually coordinate quantities to construct a linear representation of numerical magnitude, to understand place value, and to perform mental calculations. Number knowledge helps children think about mathematical problems, and its development is sensitive to children's early experiences with number (Griffin, Case, & Siegler, 1994; Siegler & Booth, 2004). Griffin et al. (1994) found that middle-income children enter kindergarten with better developed number knowledge (e.g., the ability to discriminate which of two numbers is bigger) than low-income children. Number knowledge development is linked to the amount of informal instruction children receive at home on number concepts (Saxe et al., 1987). Number knowledge is a strong predictor of arithmetic achievement in first grade, with a zero-order predictive validity correlation of .73 (Baker et al., 2002).

Number Transformations

During preschool, children acquire quantitative abilities that are relevant to learning conventional arithmetic operations. However, children have limited success in performing verbally presented calculation problems, such as story problems ("Mike had 2 pennies. Jen gave him 3 more pennies. How many pennies does he have now?") and number combinations ("How much is 2 and 3?") (e.g., Ginsburg & Russell, 1981; Levine, Jordan, & Huttenlocher, 1992). Although this might indicate a lack of calculation skill, a number of other factors could compromise the child's ability to solve story problems and number combinations. For example, children might not understand the words and syntactic structure of a problem and/or may have trouble accessing mental representations of quantities when physical referents are not provided. Levine et al. (1992) developed a "nonverbal" calculation task that eliminated these sources of difficulty. The task requires a child to reach an exact solution to a calcu-

lation problem rather than to make a judgment about the effects of the addition or subtraction transformation. Young children's success in solving nonverbal calculations depends on the ability to hold and manipulate quantitative representations in working memory as well as an understanding of number transformations (Klein & Bisanz, 2000). The ability to solve nonverbal calculation problems develops earlier than the ability to solve comparable story problems and number combinations in most children (Levine et al., 1992). Moreover, nonverbal calculation ability varies less across social classes than does the ability to solve verbal calculations (which clearly favors middle- over low-income children) (Jordan, Huttenlocher, & Levine, 1992; Jordan et al., 1994).

Estimation

There is a significant, positive relationship between estimation abilities and skill with arithmetic operations, which requires an exact knowledge of numbers (Dowker, 1997; Rubenstein, 1985). Before learning conventional arithmetic, children as young as $4\frac{1}{2}$ years old can estimate concrete set sizes and use reference points (Baroody & Gatzke, 1991). For example, they can estimate the number of dots in a set (e.g., 8) within 25% of the actual value, gauge whether a set of dots was larger or smaller than a stated reference point (e.g., "Is this group of dots more than 5 or less than 5?"), and judge whether a set fits within two reference points (e.g., "Is this group of dots less than 5, between 5 and 10, or more than 10?"). Performance varies according to set size and number of reference points (i.e., one reference point is easier than two).

Number Patterns

Recognition and use of number patterns increases skill with number combinations (Threfall & Frobisher, 1999). There are patterns within number combinations and relationships between them. For example, patterns in combinations of 6 include 3+3 = 6; therefore 4 (which is 1 more than 3)+2 (which is 1 less than three) = 6, 5+1 = 6, so 1+5 = 6, and so forth. Children with an intuitive grasp of number patterns can readily derive answers from known combinations to solve unknown ones. This ability, in turn, helps them master or become fluent with number combinations (Gray & Tall, 1994; Jordan et al., 1994). Sense of number patterns is a key component of early mathematical knowledge (Ginsburg, 1997). Children as young as 4–5 years of

age show skill in copying and extending simple number patterns (e.g., ABAB patterns), and this skill is sensitive to background variables, such as social class (Starkey, Klein, & Wakely, 2004). Good ability to discern number patterns should help children master number combinations.

Research Plan

In this study we gave our number sense battery to a large group of kindergartners four times during the school year. The study is the first part of a multiyear longitudinal project by our research team. Although we expected the various tasks to be related, we were interested to see if the tasks clustered together to form a "unitary" construct of number sense or whether more than one factor would emerge. For example, some tasks may rely more heavily on conventional knowledge whereas others may involve basic number abilities (Huttenlocher et al., 1994; Levine et al., 1992).

Longitudinal assessment with four time points allowed us to examine growth curves on the various tasks from the beginning of kindergarten, when children entered school with a wide range of experiences but with limited formal instruction, to the end of kindergarten, when all children had one year of exposure to the same math curriculum. Because early number tasks are sensitive to income level (Starkey et al., 2004), and many low-income children are at risk for developing mathematics difficulties (Jordan et al., 1994), a primary focus was on comparing performance and growth in children from low- and middle-income families. Our low-income participants were mainly African American and Latino children from urban areas. Although we anticipated that low-income kindergartners, on average, would enter school at a lower level than middle-income kindergartners (Starkey et al., 2004), we were interested in the extent to which low-income kindergartners catch up with their middle-income peers during the course of the year.

We also considered gender in our analyses. Previous work with second- and third-grade children (Jordan et al., 2003a) showed that boys had a small but consistent advantage over girls on estimation and place value tasks and that boys used concrete strategies less often than girls on number combinations. In the present investigation, we sought to determine if gender differences are evident in younger children on related tasks. Another variable we considered was reading skill. As noted earlier, children's rate of achievement in primary school math is associated with their skill in reading (e.g.,

Jordan et al., 2002). Reading skill may reflect some general level of competence or may be relevant primarily to conventional number tasks that have a basis in language (e.g., counting, story problems). Finally, we covaried age of kindergarten entry because children start kindergarten at very different ages. Holding other factors constant, we expected older children to have an advantage over younger children on our tasks.

Our statistical approach centered around two sets of analyses: conventional growth curve modeling and growth mixture modeling. Conventional growth curve modeling gives an estimate of the average level of number sense competency at a chosen time point (in this case, the end of kindergarten) as well as the average growth rate over time for all children (Hedeker, 2004; Raudenbush & Bryk, 2002). We performed conventional growth curve analyses for the different areas of number sense described previously (e.g., counting, number knowledge, number transformation, etc.). A limitation of conventional growth curve modeling is that it assumes that there is one average growth trajectory that describes the population under investigation. It may be more realistic, however, to examine whether there are distinct growth trajectory classes. For example, some children may have rapid rates of growth in number sense that level off quickly, others may have relatively steady rates of growth, while still others may have very slow rates of growth. In longitudinal studies of early reading, researchers have identified at least three unique trajectory classes (Kaplan, 2003; Leppanen, Niemi, Aunola, & Nurmi, 2004; Muthén, Khoo, Francis, & Boscardin, 2002). Distinct trajectory classes appear to emerge in number sense development as well (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004). Identification of distinct trajectory classes is important because it could lead to the finding that predictors such as age-of-entry, gender, reading proficiency, and income level do not demonstrate "one-size-fits-all" effects. Rather, the influence of these predictors may very well be different depending on the trajectory class.

To address the issue of whether there are distinct trajectory classes, we used growth mixture modeling (GMM) (Muthén, 2004) as implemented in the Mplus software program (Muthén & Muthén, 2004). The use of GMM, however, does not obviate the need to describe the sample using conventional growth curve modeling. Indeed, conventional growth curve modeling serves as an important starting point for growth mixture modeling. A technical discussion of growth mixture modeling is given in the appendix.

Method

Participants

Recruitment packages were distributed to all in-coming kindergartners in six schools in the same school district. These sites were selected for several reasons. First, each school used the *Trailblazers* kindergarten math curriculum (Teaching Integrated Mathematics and Science Curriculum, 2004), allowing us to control for the type of math instruction that the children received. Second, we chose schools whose children varied both in income status (as measured by the percentage of children receiving free/reduced lunch) and in racial/ethnic composition (as reported by parents). Two schools were located in an urban, low-income area, and one of these schools ran full-day kindergarten programs. The other four schools were located in suburban locations and ran part-day kindergarten programs. One of the urban schools was also host to the school district's Spanish–English bilingual kindergarten program, in addition to their regular kindergarten program. The full-day program differed from the part-day program in terms of special subjects rather than academic instruction (i.e., the full-day program included art classes, etc. while the part-time program did not).

Consent forms were returned for 66% of the children, resulting in an initial sample of 418 children. Seven of these children withdrew from school before the first round of data collection, leaving a final sample of 411 children, with an average age of 5 years 7 months. The background characteristics of these children are given in Table 2.

We sent surveys to families at the beginning of kindergarten to learn about the time that caregivers spent with their children on math and reading activities. The surveys were attached to the initial parental consent forms. This survey listed five math activities (e.g., counting objects, talking about numbers) and five literacy activities (e.g., reading books, saying/singing the alphabet), and asked parents to circle the frequency (rarely, sometimes, frequently) that best described the amount of time they spend on each activity with their child. Spanish-language surveys were sent to the parents of children in the bilingual program. Surveys were returned for 88% of the children (for 90% of the middle-income families and for 85% of the low-income families). Twelve returned surveys were excluded from the analysis because they contained incomplete data. We found a significant difference in the reported frequency of math and literacy activities between low-income ($M = 13.40$, $SD = 4.23$) and middle-income families

Table 2

Demographic Information of Participants by Income Status (n = 411)[a]

	Low income	Middle income	Total
Gender			
Male	68	150	218 (53 %)
Female	69	124	193 (47 %)
Race			
Minority[b]	122	110	232 (57%)
Nonminority	13	163	176 (43%)
Unknown	2	1	
Special education services			
Receiving services	8	21	29 (7%)
Not receiving services	129	253	382 (93%)
Bilingual program			
Participating	27	7	34 (8%)
Not participating	108	269	377 (92%)
Kindergarten program			
Part day	70	256	326 (79%)
Full-day	66	19	85 (21%)
Mean time 4 raw reading score (DIBELS) (*SD*)			
Letter naming fluency	32.46 (15.21)	42.97 (17.23)	39.52 (17.29)
Phoneme segmentation fluency	26.73 (18.47)	32.20 (18.29)	30.40 (18.51)
Nonsense word fluency	18.98 (13.14)	28.20 (21.61)	25.16 (19.70)

[a]Income status was determined by the participants' status in free/reduced lunch programs. "Low income" refers to students receiving free or reduced lunch, whereas "Middle income" refers to students not receiving this benefit.
[b]Minority refers to participants who self-identified as black (36.1%), Asian (5.1%), or Hispanic (15.7%) on District registration forms.

($M = 14.63$, $SD = 3.57$), with middle-income families reporting more frequent activities than low-income families, $t(349) = 2.662$, $p = .008$, $d = .32$. No significant effect was found for the gender of the child on reported frequency of literacy and numeracy activities (boys: $M = 14.21$, $SD = 3.941$; girls: $M = 14.28$, $SD = 3.711$; $t(349) = 0.167$, $p = .867$, $d = .02$).

To document the amount of children's exposure to math in kindergarten, we observed all of the kindergarten classrooms twice during the year: once in January and once in May. Five fully trained testers who had previously taught in classrooms or had experience with school observations conducted the observations. We recorded the total number of minutes spent on math activities during the kindergarten day as well as the number of math activities in each class. Observers noted the beginning and end of math activities and used a checklist based on the *Trailblazers* curriculum to indicate what content was

covered during each activity. To establish reliability, pairs of observers conducted simultaneous observations in the same classrooms during their first experience in the field. Inter-rater reliability for the training round, averaged across observers, was .96, with a range from .86 to 1.00. Following these initial classroom visits, discrepancies were discussed and resolved by mutual agreement. In addition, the observation team held two calibration meetings: once after the training round and again before the scoring of the observation data.

The average math activity lasted just over 16 min ($SD = 12.63$), and children were exposed to an average of six math activities (or activities that included math) per day, with a range of 2–12 activities within any one kindergarten day. Math activities were defined as explicit math instruction, as well as math instruction that was embedded in other content areas or daily classroom activities (e.g., integrating math when reading a story book, counting off while lining up to leave the classroom). Analyzing the data at the class level, children in the full day program participated in an average of five activities per day ($SD = 1.60$) and those in the part day program participated in an average of 7 ($SD = 2.40$) activities per day, $t(40) = 2.09$, $p = .043$, $d = .66$. Examining the data at the activity level, the mean length of an activity was 20 min ($SD = 12.99$) for the full-day program and 15 min ($SD = 12.25$) for the part-day program, $t(265) = 3.15$, $p = .002$, $d = .46$. Therefore, although children in the part-day program actually participated in more activities on average, these activities were shorter than those conducted in the full-day classrooms. As would be expected in kindergarten, the majority of activities were led by the teacher and were structured around whole-class instruction or center-based groups of children working under the supervision of the teacher or another adult. There was little variation in these patterns according to school.

Materials and Procedure

Children were assessed on the number sense battery four times during the kindergarten school year: September (Time 1), November (Time 2), February (Time 3), and April (Time 4). Children's reading skills were assessed at Time 4. Schools were tested in the same order for each 1-month testing period. Children were given the battery in school by 1 of 11 researchers who were fully trained in the testing procedures. Native or near-native speakers of Spanish administered the battery to the children in the bilingual program. Although the number sense instructions were read in English to all participants, children in the bilingual program were allowed to ask that they be clarified in Spanish and/or to answer in Spanish. If a child did not understand the task items in English, the items were also administered in Spanish. The reading measure, however, was administered to all children in English as prescribed by the school district. Children were tested individually at each of the four time points. The testing time was about 30 min per session. The number tasks were presented as games to hold children's attention. Short breaks were provided between tasks as necessary.

Number Sense Measures

Children were given the same number sense tasks at each time point. Because of the young age of the children and the 2-month period of time between assessments, we did not use alternate forms. The tasks were presented to each child in the following order: counting skills, number recognition, number knowledge, nonverbal calculation, story problems, number combinations, estimation, and number patterns. The number sense battery was shown to kindergarten teachers in all of the participating schools, who verified its relevance to the kindergarten math curriculum. Items were added or removed from the battery on the basis of teachers' suggestions.

Counting skills. The counting skills portion of the number sense battery comprised three sections: enumeration, count sequence, and counting principles. These sections were collapsed to produce one counting skills score. The total possible score on counting skills was 13 points.

Enumeration. There were four items in this section, adapted from Jordan, Levine, and Huttenlocher (1994). Children were shown five stars on a paper and asked to touch each star as they counted them. The paper was then hidden from view, and the children were then asked how many stars were on the paper (cardinality principle). The same procedure was followed using seven stars. Children were scored correct/incorrect for each question, resulting in a possible score of four points (two for counting and two for cardinality).

Count sequence. There was one item in this section. 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 they were allowed to self-correct any number that they were producing.

Counting principles. This section consisted of eight items. The task was adapted from Geary, Hoard, and

Hamson (1999) to assess children's skill at detecting violation of counting principles. It assessed children's knowledge of one to one correspondence, order irrelevance, and abstraction principles. For each item, children were presented with a set of either five or nine alternating yellow and blue dots. Children were then shown a finger puppet and told that the puppet was just learning to count. The child's task, therefore, was to tell the puppet whether he counted "OK" or "not OK." Children were given eight trials, including four correct, two incorrect, and two pseudo correct. Correct counting involved counting from left to right and counting from right to left. One pseudo error involved counting the yellow dots first and then counting the blue dots, whereas the other pseudo error involved counting the blue dots first and then the yellow dots. The incorrect counts entailed counting from left to right but counting the first dot twice. Children received a score of correct/incorrect for each trial, resulting in a possible eight points.

Number knowledge. This task was adapted from Griffin (2002). 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 either which number was bigger or which number was smaller. Last, children were shown three numbers, each of which was placed in a corner of an equilateral triangle. Children were asked to identify which number was closer to the number in the top corner of the triangle. This item assesses whether the child comprehends the relationship between numbers or is focusing on physical distance. Children were scored correct/incorrect for each question asked, for a possible eight points.

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, allowing her to manipulate chips under the box. 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 say or show how many chips were hiding under the box lid. Children had a set of chips next to their mat. After these trial rounds, four addition problems and four subtraction problems resulting in sums and minuends of seven or less were presented ($2+1$; $4+3$; $2+4$; $3+2$; $3-1$; $7-3$; $5-2$;

$6-4$). The examiner placed a quantity 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. We either added or removed chips, one at a time, and told the child how many chips were being added or removed. In order to keep the task as straightforward as possible, all addition items were given before proceeding with the subtraction items. For each item, the children were asked to say or show how many chips were left under the box. Children's incorrect performances were corrected only on the initial addition and subtraction problems.

The item was scored as correct if the child displayed the correct number of disks and/or gave the right number word. In cases where children gave both types of responses, inconsistent answers were very infrequent ($<1\%$). Children received scores of correct/incorrect for each of the addition and subtraction items, allowing a possible eight points.

Story problems. Children were given four addition and four subtraction story problems. The calculations were the same as the ones used in nonverbal calculation. The addition story problems were presented before the subtraction story problems. The addition problems were phrased as follows "Jill has two pennies. Jim gives her one more penny. How many pennies does Jill have now?," while the subtraction problems were phrased: "Mark has three cookies. Colleen takes away one of his cookies. How many cookies does Mark have now?" Children received a score of correct/incorrect for each story problem, for a possible total of eight points.

Number combinations. We gave four addition and four subtraction problems using the same calculations that were presented in Nonverbal Calculation and Story Problems. The addition problems were given before the subtraction problems. The items were phrased as follows: "How much is 2 and 1?" and "How much is 5 take away 2?" Children were scored as correct/incorrect for each number combination with the total possible score being eight points.

Estimation. This subtest was adapted from Baroody and Gatzke's (1991) estimation task. We presented children with five 15×23 cm cards with 3, 8, 15, 25, or 35 dots placed haphazardly on each card. The cards were presented to the child in the above order so that the number of dots increased each time they were shown a new card. The examiner displayed each card for as long as it took to say, "About how many dots do you see?" The child received a correct score if he or she estimated within 25% of the actual number of faces on the card. For example, in

response to the card with three dots, "3" was the only acceptable answer, whereas, in response to the card with 25 dots, any number in the range of 19–31 was scored as correct. The total possible score on Estimation was eight points.

Number patterns. The number patterns task was adapted from Starkey et al. (2004). Children were shown pictures of red (R), blue (B), and yellow (Y) beads on a string. Primary colors were chosen for ease of color distinction. In addition, a brief color-matching test was administered to each child before the first item was given in order to screen for color blindness. The tester pointed to each color bead, one at a time, and asked the child to point to the matching color at the bottom of the page. This was done for the first two items so that every color was accounted for. No children were excluded for color blindness.

For the number patterns task items, the beads were arranged in repeating patterns involving two colors (RBRBRB) and three colors (RRBYYRRBYYRR BYY); increasing patterns involving two colors (RBRBBRBBB) and three colors (RBYRRBYRRRBY); and decreasing patterns involving two colors (BBBR BBRBR) and three colors (RBYYYRBYYRBY). In each of the seven trials, one bead was left uncolored (e.g., RRBYYRRBYYRRB?Y). Children had to look at the red, blue, and yellow color choices at the bottom of each page and point to which color the uncolored bead should be in order to complete each number pattern correctly. Children's responses were scored as correct or incorrect for each pattern, for a total possible score of seven points.

Number recognition. This subtest was added to our battery at the suggestion of kindergarten teachers in the school district, who indicated that number recognition is emphasized in the kindergarten math curriculum, and consisted of four items. Children were shown the numbers 2, 8, 9, and 13 and asked to name the numbers. Children received a score of correct/incorrect for each number, for a total of four possible points.

Reading Measure

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) 6th ed. (Good & Kaminski, 2002) was given to children at the end of kindergarten. 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 min. Test–retest reliability at the end of kindergarten was .93 for letter naming fluency, .88

for phoneme segmentation fluency, and .92 for non-sense word fluency. On the basis of normative data, end of kindergarten children are considered at "some risk" if they identify fewer than 40 letters per minute (<25 high risk), fewer than 35 sounds per minute (<10 high risk), and fewer than 25 nonsense words (<15 high risk) (Good, Simmons, Kame'enui, Kaminski, & Wallis, 2002). We combined the score for each fluency area to get a total reading score. The combined measure gave us a reliable indication of general English reading proficiency.

Data Analytic Methods

As noted earlier, we used conventional growth curve modeling in the initial statistical analyses. The power of conventional growth curve modeling not-withstanding, a fundamental limitation of the method is that it assumes that the observed growth trajectories are a sample from a single finite popu-lation of individuals characterized by a single aver-age status parameter and a single average growth rate. However, it may be the case that the sample is derived from a mixture of populations, each having its own unique growth trajectory. If this is so, then conventional growth curve modeling applied to a mixture of populations will result in biased estimates of growth. Moreover, from an intervention perspec-tive, the use of conventional growth curve modeling in the presence of mixtures of populations could result in a lack of power to detect the influence of interventions on growth factors. For example, an intervention may be beneficial for children with problematic growth trajectories but irrelevant for children with positive and steep growth trajectories. Therefore, it is necessary to relax the assumption of a single population of growth and allow for the pos-sibility of mixtures of different populations.

Results

The internal reliability estimates for the full number sense battery as well as the subtests, by time point, are given in Table 3. Internal consistency for the full battery is sufficiently high across all four time points (at least ≥ .8). Reliabilities for individual subtests were somewhat lower, and thus the data should be viewed in a cautionary light. Alpha coefficients for counting and number recognition at Time 1 were particularly low and internal reliability for number patterns was low across the time points.

Although not the central purpose of this paper, it was of interest to examine the dimensionality of the number sense battery. An exploratory factor analysis

Table 3

Cronbach's Alpha for the Full Number Sense Battery and Individual Subtests by Time of Testing

	Time 1	Time 2	Time 3	Time 4
Full battery	.80	.87	.88	.82
Counting skills	.25	.39	.41	.49
Number recognition	.29	.72	.70	.63
Number knowledge	.57	.59	.62	.58
Nonverbal calculation	.51	.58	.53	.64
Story problems	.58	.71	.73	.77
Number combinations	.78	.78	.82	.85
Estimation	.42	.37	.41	.47
Number patterns	.20	.33	.31	.31

with maximum likelihood estimation and oblique rotation via PROMAX was conducted for all four time points. The results are shown in Table 4. We initially extracted two and three factors. In all cases, the three 3-factor solution resulted in a so-called Heywood case, namely a negative estimated residual variance. This problem is typically due to an over-extraction of the number of factors. The two-factor solution, however, did not result in a Heywood case.

The results in Table 4 show that across the four time points, the two-factor solution remains stable, with the first factor indicating ''basic number skills'' (counting, number recognition, number knowledge, nonverbal calculation, estimation, and number patterns) and the second factor indicating ''conventional verbal arithmetic'' (story problems and number combinations). The two-factor solution also shows adequate fit to the data as indicated by the root mean square error of approximation (RMSEA) (see lower

section of Table 4). Not surprisingly, the factors are highly correlated, but we do find that the factor correlations are relatively stable over the time periods. The factor analyses suggest a certain degree of construct validity for our number sense battery. Nevertheless, we view it as important to the main purpose of this study to examine differences in growth trajectories on the component subskills of number sense.

The findings for the baseline conventional growth curve models are presented next, followed by conventional growth curve models predicting number sense development with the addition of the covariates of gender, age of entry into kindergarten, and reading status. We then examine the added contribution of income status to number sense development over and above the first three covariates. Following the presentation of the conventional growth curve models, we provide the results of the growth mixture models for the full number sense battery as well as for a selected set of the subtests.

Conventional Growth Curve Modeling Results

The results of the baseline growth curve models are shown in Table 5. Because we were interested in end of kindergarten skills, we centered the intercept at Time 4. Therefore, intercept coefficients may be interpreted as kindergarten exit scores. As expected, the average overall number sense score and subtest scores at the end of kindergarten were significantly different from zero (e.g., the predicted average exit score in overall number sense is 38.34). As can be seen by the slope coefficients, the linear increase in

Table 4

Oblique Rotated Factor Loadings for the Number Sense Battery across Times 1–4

	Time 1		Time 2		Time 3		Time 4	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Counting skills	.977	− .067	.946	− .083	.899	− .124	− .946	− .105
Number recognition	.652	.220	.856	.011	.866	.014	.900	− .012
Number knowledge	.665	.278	.827	.092	.761	.156	.761	.177
Nonverbal calculation	.529	.339	.495	.396	.567	.349	.459	.448
Story problems	.103	.673	− .105	.927	.058	.788	− .066	.940
Number combinations	− .077	.881	.075	.724	− .025	.893	.048	.793
Estimation	.622	.216	.483	.299	.498	.199	.533	.187
Patterns	.769	− .031	.553	.265	.670	.079	.657	.076
r	.679		.692		.650		.657	
RMSEA (*p*-value)	.062 (.202)		.048 (.509)		.034 (.775)		.059 (.266)	
CI (RMSEA)	.036, 0.088		.018, 0.076		.000, 0.065		.032, 0.085	

Note. CI = confidence interval; RMSEA = root mean square error of approximation.

Table 5

Baseline Model: Growth Curve Results for Number Sense Total and Subtasks

	Number sense total	Counting skills	Number recognition	Number knowledge	Nonverbal calculation	Story problems	Number combinations	Estimation	Patterns
Intercept	38.34*	11.10*	3.58*	3.58*	4.96*	3.02*	3.43*	2.32*	4.03*
Slope	0.66*	0.19*	0.06*	0.06*	0.06	0.12*	0.14*	0.05	− 0.20
Var (intercept)	79.35*	2.50*	0.59*	0.59*	**	**	5.59*	0.68*	0.81*
Var (slope)	0.99	0.12	0.05*	0.05*	**	**	0.11	0.09	0.11
Acceleration variable	− 0.07*	0.00	− 0.01*	− 0.01*	− 0.02*	− 0.00	− 0.01	− 0.00	− 0.02*

Note. Var ( ) stands for the variance of the parameters in parentheses.
*$p < .05$. **Initial fitting of these models resulted in negative variances; therefore, models were re-estimated after fixing variance components to zero.

overall number sense was statistically significant as is the case for all subtests except nonverbal calculation, estimation, and number patterns. In general, the rate of acceleration slowed over time as evidenced by the negative and statistically significant acceleration parameters for overall number sense, number recognition, number knowledge, nonverbal calculation, and number patterns.

Table 6 presents the results of the conventional growth curve models with the addition of gender, grand-mean centered kindergarten start age, and reading proficiency (z score). The coefficients given in the top portion of the table indicate the predicted values for girls of an average kindergarten start age (5 years 7 months) with an average reading score. Focusing on the effects of the covariates, we found gender differences at the end of kindergarten favo-

ring males for overall number sense, counting skills, number knowledge, nonverbal calculation, estimation, and patterns. For example, while girls of average start age and reading scores were predicted to score 10.92 on the counting skills subtest at kindergarten exit, boys of average start age and reading scores were predicted to score 0.35 points higher on this measure. With the exception of overall number sense, boys did not differ from girls in the linear rate of growth over time, and there were no gender differences in acceleration over time. With regard to kindergarten start age, age of entry into kindergarten measured in months was positively associated with end of kindergarten performance on overall number sense, number recognition, number knowledge, nonverbal calculation, story problems, and number combinations. For every month older that a child

Table 6

Model 1: Growth Curve Results for Number Sense Total and Subtasks with Effects of Gender, Mean-Centered Kindergarten Start Age, and Time 4 Reading z Score

	Number sense total	Counting skills	Number recognition	Number knowledge	Nonverbal calculation	Story problems	Number combinations	Estimation	Patterns
Intercept	37.21*	10.92*	3.54*	5.92*	4.67*	2.83*	3.23*	2.17*	3.91*
Slope	0.36	0.15*	0.04	0.01	− 0.01	0.10	0.04	0.02	− 0.08
Var (intercept)	49.75*	1.92*	0.38*	1.26*	**	**	3.78*	0.62*	0.79*
Var (slope)	1.49	0.06	0.04	0.05	**	**	0.07	0.11*	0.12
Acceleration variable	− 0.11*	− 0.00	− 0.01*	− 0.02*	− 0.03*	− 0.00	− 0.02*	− 0.00	− 0.03*
Intercept on male	2.77*	0.35*	0.09	0.30*	0.59*	0.41	0.44	0.31*	0.27*
Slope on male	0.68*	0.10	0.05	0.00	0.12	0.07	0.19	0.05	0.11
Acceleration on male	0.07	0.01	0.01	0.00	0.01	0.00	0.03	0.00	0.02
Intercept on K start age	0.38*	0.03	0.03*	0.06*	0.06*	0.06*	0.08*	0.03	0.03
Slope on K start age	0.01	0.00	0.00	− 0.01	0.00	− 0.01	0.01	0.01	− 0.01
Acceleration on K start age	0.00	0.00	0.00	− 0.00	0.00	− 0.00	0.00	0.00	− 0.00
Intercept on reading score	5.17*	0.56*	0.41*	0.70*	0.76*	1.03*	1.22*	0.29*	0.22*
Slope on reading score	0.06	0.03	− 0.02	− 0.05	0.03	0.05	0.04	0.03	− 0.05
Acceleration on reading score	− 0.01	0.00	0.00	− 0.01	− 0.00	− 0.00	− 0.00	0.01	− 0.01

Note. Var ( ) stands for the variance of the parameters in parentheses.
*$p < .05$. **Initial fitting of these models resulted in negative variances; therefore, models were re-estimated after fixing variance components to zero.

Table 7

Model 2: Growth Curve Results for Number Sense Total and Subtasks with Effects of Gender, Mean-Centered Kindergarten Start Age, Time 4 Reading z Score, and Income Status

	Number sense total	Counting skills	Number recognition	Number knowledge	Nonverbal calculation	Story problems	Number combinations	Estimation	Patterns
Intercept	39.35*	11.10*	3.62*	6.20*	4.89*	3.31*	3.75*	2.29*	4.12*
Slope	0.53*	0.18*	0.01	0.05	− 0.01	0.19*	0.05	0.04	− 0.10
Var (intercept)	43.50*	1.9*	0.38*	1.01*	**	**	3.23*	0.61*	0.80*
Var (slope)	1.62	0.07	0.04	0.02	**	**	0.04	0.11*	0.14
Acceleration variable	− 0.10*	− 0.00	− 0.02*	− 0.02	− 0.03*	0.00	− 0.03*	− 0.00	− 0.03*
Intercept on male	2.37*	0.32	0.07	0.24	0.55*	0.32	0.34	0.29*	0.23
Slope on male	0.65*	0.09	0.05	− 0.00	0.12	0.06	0.19	0.05	0.11
Acceleration on male	0.07	0.01	0.01	0.00	0.01	0.00	0.03	0.00	0.02
Intercept on K start age	0.36*	0.03	0.03*	0.05*	0.05*	0.05*	0.08*	0.02	0.03
Slope on K start age	0.00	0.00	0.00	− 0.01	0.00	− 0.01	0.01	0.01	− 0.01
Acceleration on K start age	0.00	0.00	0.00	− 0.00	0.00	− 0.00	0.00	0.00	− 0.00
Intercept on reading score	4.47*	0.50*	0.38*	0.60*	0.68*	0.87*	1.05*	0.25*	0.16*
Slope on reading score	0.01	0.02	− 0.01	− 0.07	0.03	0.02	0.04	0.02	− 0.04
Acceleration on reading score	− 0.02	0.00	0.00	− 0.01	− 0.00	− 0.01	− 0.00	0.00	− 0.01
Intercept on LOW INCOME	− 5.83*	− 0.49*	− 0.24*	− 0.79*	− 0.61*	− 1.31*	− 1.44*	− 0.34*	− 0.58*
Slope on LOW INCOME	− 0.46	− 0.07	0.09	− 0.11	0.01	− 0.25*	− 0.04	− 0.05	0.06
Acceleration on LOW INCOME	− 0.01	− 0.00	0.00	− 0.02	0.01	− 0.02	0.01	− 0.01	0.01

Note. Var ( ) stands for the variance of the parameters in parentheses.
*$p < .05$. **Initial fitting of these models resulted in negative variances; therefore, models were re-fit after fixing variance components to zero.

entered kindergarten, her total battery score increased by 0.38 points. Age of entry showed no significant relationship to linear growth or acceleration on any of the subtests. In other words, older children began kindergarten slightly more advanced in number sense skills than their younger classmates, and this difference persisted throughout the year. With regard to reading proficiency, we found that better performance on the reading test was associated with positive and statistically significant performance at the end of kindergarten on overall number sense and all subtests. Reading proficiency was not significantly associated with a linear trend or with acceleration over time.

Table 7 presents the growth curve results when income status is added to the model. The trajectories by task, income level, and gender are plotted in Figure 1. Children in the middle-income group performed significantly better than children from the low-income group at the end of kindergarten on the overall number sense battery and on all of the subtests. Furthermore, a significant effect of income status on the slope suggests that low-income children experienced less growth on story problems throughout the kindergarten year than middle-income children. No acceleration effects of income were revealed. The addition of the income status did not appreciably change the findings of previous covariates entered into the model.

Growth Mixture Modeling Results

As noted earlier, a goal of this paper was to explore the extent to which there exist underlying latent classes of growth trajectories in number sense and subtest performance, and how these latent growth trajectory classes might be differentially predicted by the background variables of gender, kindergarten start age, reading proficiency, and income status. In addition to total battery performance, we selected the nonverbal calculation, story problems, and number combinations subtests for further analysis for two reasons. First, our growth curve analyses showed divergent trajectories of growth on these tasks based on gender and income (see Figure 1), which suggested that a growth mixture approach to the tasks would provide useful information on how gender and income status may play a role in defining a population of children who are at risk for mathematics difficulties. Moreover, these three tasks assessed manipulation of identical sets of calculations in three different contexts.

The results of the growth mixture modeling are displayed in Table 8 for the total number sense and for the three selected subtests. Frequencies of children placed in each class by demographic group and task are presented in Table 9. The number of children assigned to classes (369) is slightly smaller than the
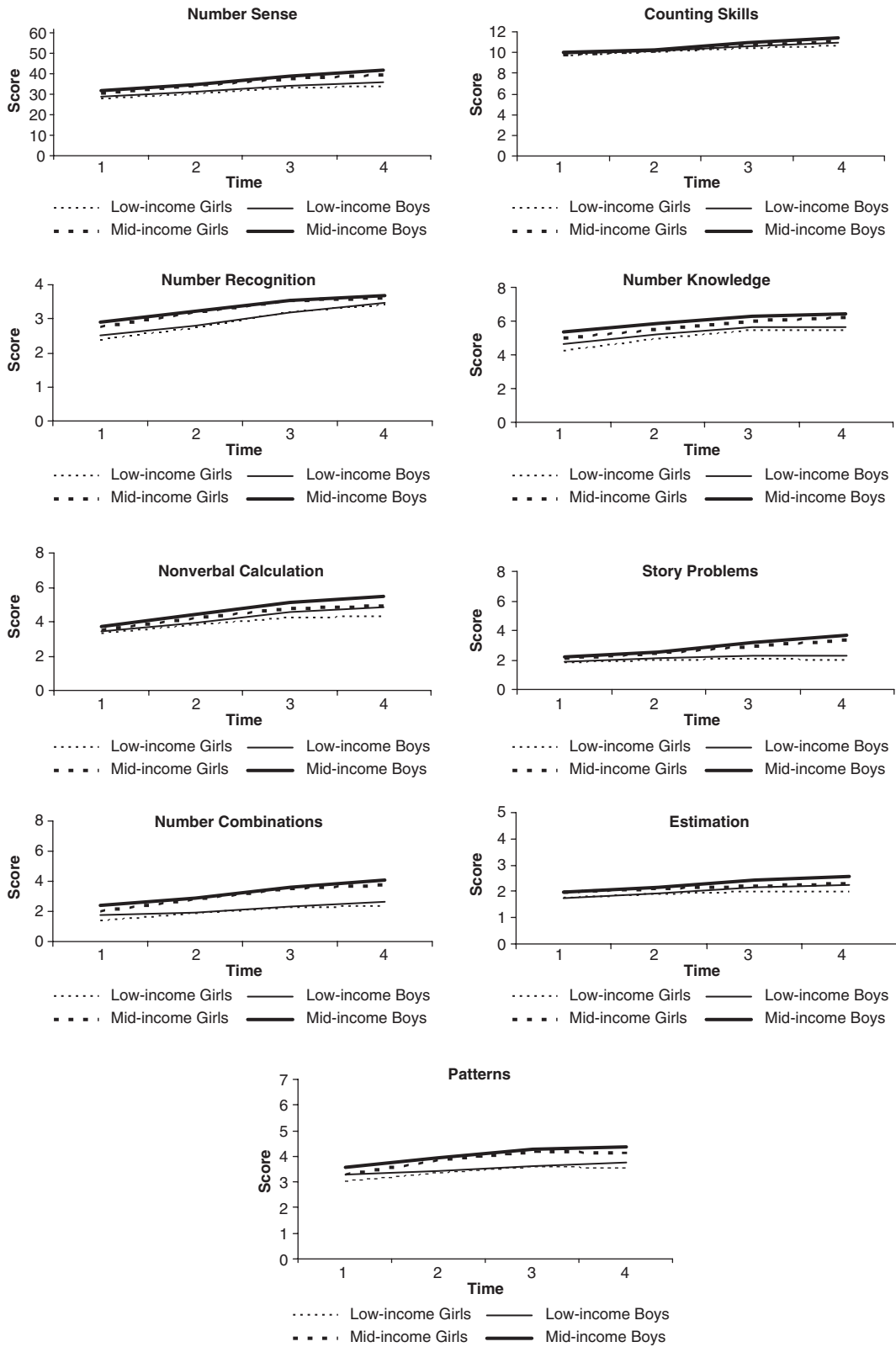
Figure 1. Fitted trajectories controlling for grand-mean centered kindergarten start age and time 4 reading score for number sense battery and subtasks by gender and income status.

Table 8

Results of Three-Class Growth Mixture Model with Effects of Gender, Grand Mean-Centered Kindergarten Start Age, Reading Score at Exit (z-score), and Income Status Number Sense Battery and Subtest Scores

	Number sense battery			Nonverbal calculation			Story problems			Number combinations		
	Class#1 Low/flat	Class#2 Avg/mod	Class#3 High/mod	Class#1 Low/flat	Class#2 High/steep	Class#3 High/mod	Class#1 Low/flat	Class#2 High/mod	Class#3 High/flat	Class#1 low/flat	Class#2 high/steep	Class#3 high/mod
Coefficient												
I	28.92*	38.20*	49.88*	3.49*	5.75*	6.67*	2.37*	0.96	6.03*	2.14*	6.16*	6.46*
S	−0.19	0.72	0.92	−0.51	0.46*	0.36	0.22	−1.10*	0.29	0.06	1.03	−0.11
Q	−0.05	0.02	−0.12	−0.10	0.03	0.02	0.01	−0.25*	−0.01	−0.01	0.04	−0.07*
I on MALE	1.54	0.50	−0.20	0.11	0.10*	−0.55	−0.15	4.22*	0.32	−0.38*	−0.64	0.28
S on MALE	0.98	0.72	0.16	−0.03	0.09	0.00	0.03	−0.08	−0.39	0.08	0.66	0.12
Q on MALE	0.08	0.13	0.04	−0.00	−0.02	0.01	0.003	0.30*	−0.03	0.02	0.14*	0.02
I on K start age	0.81*	0.89*	0.45*	−0.04	−0.10*	0.07	0.02	−0.08	0.09	0.06*	0.32*	−0.05
S on K start age	0.01	0.12	−0.05	−0.03	−0.06	0.00	0.01	−0.13*	0.02	0.02	0.19*	0.01
Q on K start age	−0.00	0.01	0.00	−0.00	−0.01	0.01	0.00	−0.02*	0.00	0.00	0.02*	0.01
I on reading	2.73*	−0.51	−0.18	0.55*	0.54*	0.16	0.41*	0.27	0.24	0.37*	0.21	0.20
S on reading	0.31	−0.46	−0.12	0.04	−0.30*	−0.02	0.06	−0.20	−0.10	0.10	0.20	−0.03
Q on reading	0.04	−0.02	0.00	0.00	−0.05*	−0.01	0.00	−0.01	−0.02	0.01	0.05	−0.00
I on LOW INCOME	1.02	−2.85	−2.00	−0.23	0.21	−0.62	−0.68*	1.43	−0.86	−0.40*	0.99	1.14*
S on LOW INCOME	0.20	−0.59	−0.38	0.21	0.10	0.09	−0.29*	0.20	0.28	−0.07	0.22	−0.11
Q on LOW INCOME	0.01	−0.08	−0.05	0.05*	0.03	−0.03	−0.03	−0.08	0.05	−0.00	−0.02	−0.05
Final class proportions	0.29	0.40	0.31	0.44	0.30	0.26	0.70	0.14	0.16	0.61	.13	0.26

Note. I = intercept; S = linear slope; Q = quadratic.
*$p < .05$.

Table 9

Frequencies of Children Placed in Each Class, by Demographic Group, for Number Sense Battery, Nonverbal Calculation, Story Problems, and Number Combinations (n = 369)

	Class #1	Class #2	Class #3	Total
Number Sense Battery	Low/flat 102 children	Avg/mod 155 children	High/mod 112 children	
Gender				
Male	48 (47%)	77 (50%)	68 (61%)	193 (52%)
Female	54 (53%)	78 (50%)	44 (39%)	176 (48%)
Income status				
Mid-income	30 (29%)	121 (78%)	97 (87%)	248 (67%)
Low-income	72 (71%)	34 (22%)	15 (13%)	121 (33%)
Mean age at kindergarten entry (years)	5.75 (0.33)	5.50 (0.25)	5.58 (0.25)	5.58 (0.33)
Mean letter naming fluency score	28.82 (14.80)	37.56 (14.97)	52.12 (14.72)	39.57 (17.35)
Mean phoneme segmentation score	20.17 (16.83)	30.79 (18.76)	39.66 (14.14)	30.54 (18.45)
Mean nonsense word fluency score	14.78 (11.65)	21.40 (13.15)	40.20 (23.99)	25.28 (19.69)
Nonverbal calculation	Low/flat 91 children	High/steep 148 children	High/mod 168 children	
Gender				
Male	38 (42%)	74 (50%)	102 (61%)	214 (53%)
Female	53 (58%)	74 (50%)	66 (39%)	193 (47%)
Income status				
Mid-income	44 (48%)	99 (67%)	128 (76%)	271 (67%)
Low-income	47 (52%)	49 (33%)	40 (24%)	136 (33%)
Mean age at kindergarten entry (years)	5.50 (0.25)	5.58 (0.33)	5.67 (0.33)	5.58 (0.33)
Mean letter naming fluency score	29.57 (16.28)	39.31 (16.67)	45.40 (15.83)	39.60 (17.31)
Mean phoneme segmentation score	22.72 (18.76)	31.40 (19.52)	33.95 (16.55)	30.47 (18.49)
Mean nonsense word fluency score	14.54 (12.34)	25.57 (18.40)	30.89 (21.63)	25.57 (19.70)
Story problems	Low/flat 260 children	High/mod 54 children	High/flat 55 children	
Gender				
Male	121 (46%)	53 (98%)	19 (35%)	193 (52%)
Female	139 (54%)	1 (2%)	36 (65%)	176 (48%)
Income status				
Mid-income	148 (57%)	50 (93%)	50 (91%)	248 (67%)
Low-income	112 (43%)	4 (7%)	5 (9%)	121 (33%)
Mean age at kindergarten entry (years)	5.58 (0.25)	5.67 (0.33)	5.67 (0.33)	5.58 (0.33)
Mean letter naming fluency score	35.33 (15.83)	48.54 (16.81)	50.78 (16.64)	39.57 (17.35)
Mean phoneme segmentation score	26.65 (19.42)	31.82 (15.94)	38.93 (14.12)	30.54 (18.45)
Mean nonsense word fluency score	19.10 (13.32)	25.98 (16.49)	39.26 (25.69)	25.28 (19.69)
Number combinations	Low/flat 223 children	High/steep 50 children	High/mod 96 children	
Gender				
Male	109 (49%)	27 (54%)	57 (59%)	193 (52%)
Female	114 (51%)	23 (46%)	39 (41%)	176 (48%)
Income status				
Mid-income	119 (53%)	38 (76%)	91 (95%)	248 (67%)
Low-income	104 (47%)	12 (24%)	5 (5%)	121 (33%)
Mean age at kindergarten entry (years)	5.58 (0.33)	5.50 (0.25)	5.75 (0.25)	5.58 (0.33)
Mean letter naming fluency score	34.29 (15.65)	39.42 (16.70)	51.90 (15.17)	39.57 (17.35)
Mean phoneme segmentation score	26.65 (19.42)	31.82 (15.94)	38.93 (14.12)	30.54 (18.42)
Mean nonsense word fluency score	19.10 (13.32)	25.98 (16.49)	39.26 (25.69)	25.28 (19.69)

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

sample size of 411 children because missing data prevented some children from being assigned to a particular trajectory class.

Considering the total number sense battery, we found three distinct classes that are labeled according to their kindergarten exit level relative to the average (low, average, or high) as well as their growth rate (flat, moderate, or steep growth). We characterized these three classes as "low/flat", "avg/mod," and "high/mod" based on the shapes
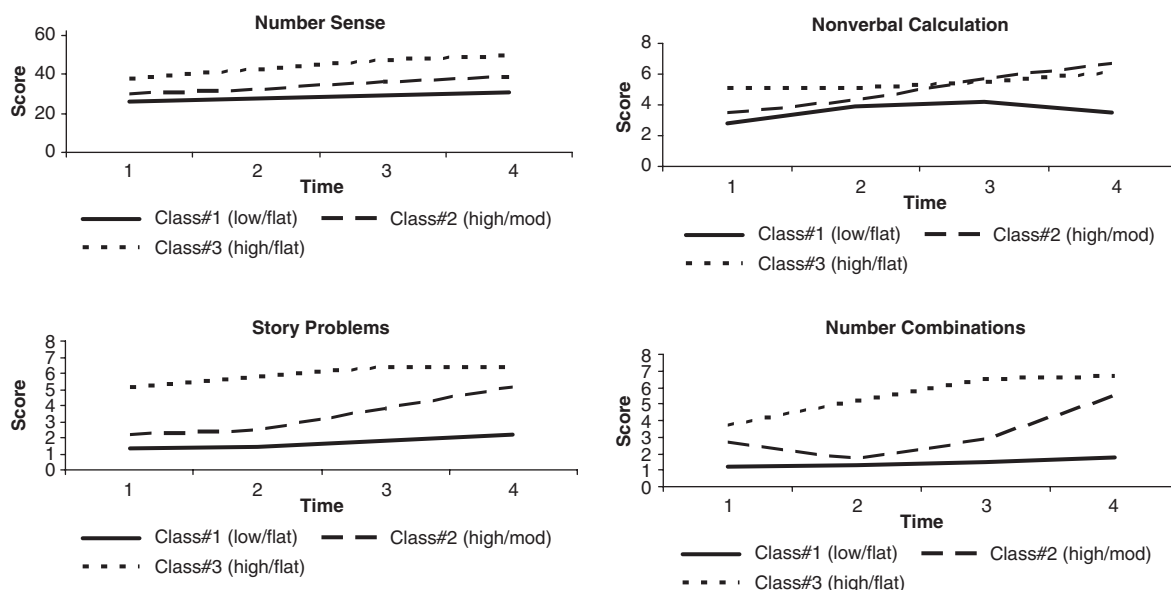
Figure 2. Fitted trajectories controlling for gender, grand-mean centered kindergarten start age, and time 4 reading score by growth mixture class.

of the trajectories (see Figure 2). The bottom of Table 8 shows that approximately 29%, 40%, and 31% of the sample fell into these three classes, respectively. Class labels and latent class proportions for the remaining subtests are also displayed in Table 8. The trajectories for total number sense as well as for the three subtests are plotted by class in Figure 2.

Turning to the influence of the predictor variables within each class for *total number sense*, there were no gender or income differences across the three trajectory classes. There was a positive and statistically significant association between the age of entry into kindergarten and exit number sense score across the three classes of total number sense, meaning that children who are older when they enter kindergarten demonstrate significantly better exit number sense scores than their younger counterparts. We found that reading proficiency was significantly associated with total number sense score at the end of kindergarten, but only for children in the low/flat class. In other words, children in the low/flat class who were better readers ended kindergarten with better number sense performance. This suggests that skill development in reading may be beneficial for low-performing children, but has no statistically significant effect for children already performing better in math.

We also found three distinct classes with respect to the *nonverbal calculation* task: "low/flat", "high/steep," and "high/mod." Approximately 44%, 30%, and 26% of our sample falls into these classes, respectively. Although there were two classes of chil-

dren who did not demonstrate statistically significant growth in nonverbal calculation in kindergarten, there was also a group of children (high/steep) whose skills in this area grew throughout kindergarten, allowing them to exit with above-average scores on this task.

The only gender difference occurred on the nonverbal calculation exit score for the high/steep class, but this difference of 0.10 points on the subtask score was minimal. Likewise, the negative effect of kindergarten start age on task performance for children in the class was not large: for every month older that a child began kindergarten, that child scored an average of 0.10 points lower on nonverbal calculation at the end of the year. The more interesting findings occurred in reading. Reading score at kindergarten exit is significantly and positively associated with the nonverbal calculation performance of children in the low/flat and high/steep classes. Furthermore, the children in the high/steep class received an additional benefit of reading on nonverbal calculation growth and acceleration. In sum, reading ability played a significant role in the growth and exit performance of nonverbal calculation skills, but only for children in the high/steep class.

Analysis of children's performance on *story problems* also revealed three distinct classes: low/flat, high/mod, and high/flat. Of the children in the sample, 70%, 14%, and 16% of the children fell into these three classes, respectively. As can be seen from Table 5, gender appeared to have a very large and significant effect on story problems skills at the end

of kindergarten for children in the high/mod class. Upon further investigation, however, we found that this class consisted of 53 boys and only 1 girl; therefore, the effect of gender was overestimated. Kindergarten start age had a small but statistically significant effect on growth and acceleration in story problems performance, but only for the children in the high/mod class. Reading showed significant effects only on story problems exit score for the low/flat class alone. Income status, however, had a more notable effect on this task: children in the low/flat class from low-income households scored an average 0.68 points lower at the end of kindergarten and grew an average of 0.29 points slower per month in story problems skill throughout kindergarten than did children from middle-income households in the same class.

On *number combinations*, children fell into three classes: low/flat, high/steep, and high/mod. The distribution of children into these classes was 61%, 13%, and 26%, respectively. Significant linear growth was not observed for any class; there was a four-point discrepancy on exit scores between the low/flat class on the one hand and the high/steep and high/mod classes on the other, as can be seen by comparing the data presented in the first line of Table 8. Gender had an effect on number combinations score for the low/flat class alone. Among these children, boys scored an average 0.38 points lower than girls at the end of the kindergarten year. Among the high/steep class, boys experienced slightly faster growth during the year. Age of kindergarten entry seemed particularly relevant for children in the high/steep class: children scored an average of 0.32 points higher on number combinations for every month later they enrolled in kindergarten. The children in this class also gained an additional 0.19 points per kindergarten month and an acceleration of 0.02 points per month for every month later they started kindergarten. Kindergarten start age also had a small but positive effect on number combinations score for children in the low/flat class. Reading, on the other hand, affected only the exit score in number combinations for children in the low/flat class. Income had different effects on kindergarten exit score depending on class. Children from low-income households in the low/flat class, for example, scored 0.40 points lower on number combinations at the end of the year than did children from mid-income households in the same class. Low-income children in the high/mod class, however, scored 1.14 points higher at exit than did middle-income children in the same class.

Table 10

Odds of Class Assignment (with Class #3 as Comparison) for Number Sense Battery, Nonverbal Calculation, Story Problems, and Number Combinations

Number Sense Battery[a]
Class #1 (low/flat)		Class #2 (avg/mod)	
MALE	0.33	MALE	0.48
K start age	1.32*	K start age	1.01
Time 4 reading score	0.07*	Time 4 reading score	0.25*
LOW INCOME	17.29*	LOW INCOME	1.36

Nonverbal Calculation[b]
Class #1 (low/flat)		Class #2 (high/steep)	
MALE	0.49	MALE	0.85
K start age	0.87	K start age	1.04
Time 4 reading score	0.47	Time 4 reading score	0.50*
LOW INCOME	1.30	LOW INCOME	0.39

Story Problems[c]
Class #1 (low/flat)		Class #2 (high/mod)	
MALE	1.09	MALE	169.02[e]
K start age	0.92	K start age	1.02
Time 4 reading score	0.43*	Time 4 reading score	1.60
LOW INCOME	4.39*	LOW INCOME	0.70

Number Combinations[d]
Class #1 (low/flat)		Class #2 (high/steep)	
MALE	0.46*	MALE	0.69
K start age	0.90	K start age	0.77*
Time 4 reading score	0.30*	Time 4 reading score	0.50*
LOW INCOME	10.49*	LOW INCOME	4.44*

[a]Comparisons are made with Number Sense Battery Class #3 (high/moderate).
[b]Comparisons are made with Nonverbal Calculation Class #3 (high/moderate).
[c]Comparisons are made with Story Problems Class #3 (high/flat).
[d]Comparisons are made with Number Combinations Class #3 (high/moderate).
[e]Story problems, Class #2 consisted of 53 boys and 1 girl; therefore, the effect of gender is overestimate.
*$p < .05$.

Multinomial Logistic Regression Analyses

An important additional analysis derived from the application of growth mixture modeling is the prediction of class membership as a function of the background variables via multinomial logistic regression. Multinomial logistic regression provides the odds of a child being assigned to one growth trajectory class versus another based on whether the child is male (or female), low income (or middle income), a good reader (or a poor reader), or started kindergarten earlier (or later). Multinomial logistic regression coefficients were estimated within the MPlus software program, and the odds ratios are displayed in Table 10. For these analyses, Class # 3 for the total number sense battery and subtests was used as the comparison group (see Table 8).

Examining the results for *total number sense*, we found that children who were older when they entered kindergarten, had lower reading proficiency and were of low-income status had greater odds of being in the low/flat class versus the high/mod class. For example, the kindergarten start age coefficient can be interpreted as indicating that for every month older a child began kindergarten that child had 1.32 times greater odds, or was approximately one-third times more likely, of falling into the low flat versus the high/mod class. The low-income coefficient indicates that children of low-income status were 17.29 times more likely to be assigned to the low/flat class rather than the high/mod class when compared with their middle-income peers. In contrast, children with better reading scores had lower odds of being in both the low/mod or avg/mod class relative to being in the high/mod class. Reading skill was the only variable that differentiated the avg/mod class from the high/mod class in number sense, but kindergarten start age, reading skill, and income status differentiated the low/flat class from the high/mod class. In other words, better reading performance will increase a child's odds of a high number sense exit score whether or not the child is in the avg/mod or the low/flat class. Low-income status, on the other hand, gives a child greater odds of having low end-of-kindergarten number sense scores.

Looking at the coefficients for the *nonverbal calculation* task, we found that children with lower kindergarten exit scores in reading had lower odds of being in the high/steep versus the high/mod class. Results from the *story problems* task showed that poorer readers had lower odds of being in the high/flat class versus the low/flat class. Low-income children had greater odds of being in the low/flat class versus the high/flat class. (The significant gender difference on story problems [between the high/mod and the high/flat class] must be interpreted cautiously, as the high/mod class contained only one girl.)

The results from the *number combinations* task showed that low-income girls and poorer readers had higher odds of being in the low/flat versus the high/mod class. Children who entered kindergarten at an older age, children with lower reading scores, and middle-income children had lower log-odds of being in the high/steep versus the high/mod class.

In sum, low income status and poor reading skill consistently increase children's odds of starting and ending kindergarten at a low level on the number tasks, with the exception of nonverbal calculation. It should be noted, however, that the effects for income are much greater than the effects for other predictors.

Discussion

We tracked the development of number sense in 411 children over four time points during kindergarten. Children's performance and growth were examined while controlling for background variables of income level, gender, age, and reading skill. We looked at children's level of performance at the end of kindergarten as well as their rates of growth from the beginning to the end of kindergarten. All children received essentially the same math curriculum throughout the school year. There did not appear to be large differences between classes or schools in terms of math exposure. Although our number sense battery showed good reliability across the various time points ($>.8$), individual subtests were less reliable and should be interpreted with caution.

Two-Dimensional Model of Number Sense

A factor analysis addressed the extent to which our number sense tasks cluster together. Our battery was two dimensional across all time points: the first dimension involving basic number skills (counting, number recognition, number knowledge, nonverbal calculation, estimation, and number patterns) and the second dimension involving conventional arithmetic (story problems and number combinations). This two-dimensional model corresponds with quantity discrimination and verbal sequential factors of mathematics proficiency reported by Okamoto and Case (1996) using similar methods of analysis. The model is also in keeping with Berch's (2005) discussion of the lower and higher order features of number sense. Lower order number sense is characterized by elementary knowledge about quantity, such as counting and comparing numerical magnitudes (Dehaene, 2001), whereas secondary skills that result from conventional educational activities characterize higher order number sense. Nonverbal calculation involves a more basic arithmetic ability, one that emerges after the approximate skills of infancy but before the conventional skills of arithmetic (Huttenlocher et al., 1994). Likewise, counting skills, estimation, and knowledge of number magnitudes underpin school arithmetic (Aunola et al., 2004; Geary, 1995).

Development of Number Sense

Growth curve analyses revealed significant linear growth throughout kindergarten in most areas of number sense. We found three distinct growth classes for our overall battery as well as for the se-

lected calculation subtests (nonverbal calculation, story problems, and number combinations). Although the characteristics of growth classes varied somewhat between tasks, we generally found a group of children who ended kindergarten at a low level with flat growth, another group of children who ended kindergarten at a high level with moderate to steep growth, and a final group of children who ended kindergarten at an average level with moderate growth. Previous longitudinal work with a sample of Finnish children (Aunola et al., 2004) identified two classes of growth in math achievement, namely low performers and high performers. Our data, using a more heterogeneous sample, further captured a class that starts kindergarten at a relatively low level but makes moderate to fast growth during the year. It is likely that our multi-ethnic U.S. sample had more diversity in terms of preschool experiences than the Finnish sample, which could account for the identification of a third class in our growth mixture results. Identification of these unique growth classes is important for differentiating children in obvious need of intervention (e.g., low/flat) from those who are likely to make progress with minimal or no special help (e.g., avg/mod).

Social Class

A primary goal of our study was to examine number sense development in children from low-income households. Although social class differences in number-related skills have been documented (e.g., Ginsburg & Russell, 1981; Jordan et al., 1994; Starkey et al., 2004), few investigations have studied young children longitudinally with multiple time points. Holding other background variables constant, low-income children in the present study fared poorly on most of our number sense tasks relative to their middle-income peers. Yet children in both income groups progressed at similar rates on many tasks. A notable exception was on story problems, where low-income kindergartners showed almost no growth from the beginning to the end of the year, even when we accounted for reading proficiency. Story problems are a conventional task that requires children to perform arithmetic calculations embedded in sentences. Story problems are demanding not only in terms of basic arithmetic but also in terms of language and auditory attention (Fuchs et al., in press; Levine et al., 1992). Although many kindergartners fell into the low-flat growth class for story problems, a low-income kindergartner was four times more likely to be in this class than his or her middle-income peers. A similar pattern was observed on number combinations, where a low-income child was much more likely to fall in the low/flat class than a middle-income child. Interestingly, both income groups made comparable progress when the same calculations were presented in a nonverbal context with visual referents. The findings corroborate those of Huttenlocher and colleagues (Huttenlocher et al., 1994; Levine et al., 1992), which revealed that nonverbal calculation abilities are less sensitive to social class than are conventional story problems. Our divergent story problem trajectories show that the income gap widens in kindergarten, despite comparable math instruction, and low-income children are likely to enter first grade with a genuine disadvantage in problem solving. It should be noted that story problems and number combinations were not introduced into the curriculum until relatively late in the year. Therefore, a possible scenario is that middle-income children are acquiring these skills outside the classroom while low-income children are not. In other words, it is not necessarily the case that low-income children are unresponsive to instruction in these areas. It should also be noted that that there may have been classroom differences between low- and middle-income children that were not revealed in the present study. For example, many of our low-income children attended a full-time kindergarten program (with the addition of special subjects, such as art) whereas all of our middle-income children attended a part-time program. In any case, because math becomes more dependent on verbal skills in primary school and difficulties in story problems and arithmetic combinations are defining characteristics of math difficulties (Jordan et al., 2002), kindergarten or early first grade may be an important time to intervene with high-risk children.

Gender

Overall, our findings suggest that gender differences in math emerge as early as kindergarten. There were small, but statistically reliable, gender effects on kindergarten level performance on overall number sense, nonverbal calculation, and estimation. In each case, boys showed an advantage over girls, and the findings held above and beyond income level, age, and reading ability. There is also a greater percentage of boys in high performing classes on overall number sense (61% vs. 39%) and on nonverbal calculation (61% vs. 39%) when background variables are controlled for. A gender effect was also reported in previous work with school-age children (Jordan et al., 2003a) where third-grade boys performed better

than girls on tasks assessing place value, estimation (i.e., approximate arithmetic), and mental computation. Although Aunola et al., (2004) did not find gender differences in level of performance, they found that boys achieved at a faster rate in primary-school math, especially among high performing classes of children. We did not see gender differences in growth rates in kindergarten. Although our gender effects may be small in practical terms, they are consistent and could help explain some of the gender differences in math problem solving that are present in later years (Hyde, Fennema, & Lamon, 1990).

Note that the skills on which boys seem to have an advantage (e.g., nonverbal calculation) involve spatial reasoning as well as some manipulation of quantities or numerical magnitudes (Baroody & Gatzke, 1991; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). The finding raises the question of whether there are cognitive explanations for gender differences in early math (Geary, 1998). Interestingly, a consistent finding in the literature on spatial cognition is that males have an advantage over females in spatial cognition when it is measured by mental rotation tasks or by spatial judgments about moving objects (Casey, Nuttal, & Pezaris, 1997; Royer, Tronsky, Chan, Jackson, & Marchant, 1999). Of course, there may also be explanations related to socialization and motivation such that boys are encouraged more than girls on number tasks, even at an early age (Aunola et al., 2004; Eccles, Janus, & Harold, 1990).

Reading

Reading proficiency at the end of kindergarten was a significant predictor of performance on all number sense measures (but not of rate of growth) even when we considered other background variables. Leppanen et al. (2004) also reported a relationship between early reading and number skills and Jordan et al. (2002) found that children with mathematics difficulties who are good readers make better progress in math achievement than children with mathematics difficulties who are poor readers. Recall that we assessed reading with a standard measure of word-level fluency skills, that is, letter naming, nonsense word reading, and phoneme segmentation. It is possible that the kindergarten reading score reflected some kind of general verbal cognitive ability. Supporting this notion, Jordan and Hanich (2003) report that elementary school children with both reading and math difficulties have lower IQ scores than do children with mathematics difficulties only or children with reading difficulties only. However, the data may also reflect a relationship between

school-related skills and early home experiences (Saxe et al., 1987). In any case, reading difficulties are a correlate of math difficulties (Leppanen et al., 2004) and should be viewed as a risk factor.

Age

As expected, age was a significant predictor of overall number sense performance at the end of kindergarten. Older children began kindergarten slightly more advanced than their younger peers and this difference held throughout the school year. Of course, the age effects are small in practical terms, and it remains to be seen whether age will predict math achievement in subsequent grades. Reading researchers have found that although older kindergartners have stronger emergent literacy skills than younger kindergartners, age is not a good predictor of reading achievement at the end of the first and second grades (Crone & Whitehurst, 1999).

The present kindergarten study is the initial phase of a longitudinal investigation of children's mathematics from kindergarten through third grade. Remaining questions include the extent to which kindergarten number sense predicts achievement in primary school, the extent to which growth classes in kindergarten predict growth classes in elementary school, and the extent to which income, gender, and age effects hold, become greater, or attenuate in elementary school. We will also continue to look at children's reading development as well as the role of general cognitive abilities in children's rate of achievement in math.

References

Ashcraft, M. H. (2002). Math anxiety: Personal educational and cognitive consequences. *Current Directions in Psychological Science, 11,* 181–185.

Aunola, K., Leskinen, E., Lerkkanen, M., & Nurmi, (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96,* 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.* (Technical Report No. 0305). Eugene, OR: Pacific Institutes for Research.

Baroody, A. J. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Education, 18,* 141–157.

Baroody, A. J., & Gatzke, M. R. (1991). The estimation of set size by potentially gifted kindergarten-age children. *Journal for Research in Mathematics Education, 22,* 59–68.

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

Briars, D., & Siegler, R. S. (1984). A featural analysis of preschooler's counting knowledge. *Developmental Psychology, 20*, 607–618.

Case, R. (1998). *A psychological model of number sense*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Case, R., & Griffin, S. (1990). Child cognitive development: The role of central conceptual structures in the development of scientific and social thought. In E. A. Hauert (Ed.), *Developmental psychology: Cognitive, perceptuomotor, and neurological perspectives* (pp. 193–230). North-Holland: Elsevier.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology, 33*, 669–680.

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling in the social and behavioral sciences* (pp. 81–110). San Francisco: Jossey-Bass.

Crone, D. A., & Whitehurst, G. J. (1999). Age and schooling effects on emergent literacy and early reading skills. *Journal of Educational Psychology, 91*(4), 604–614.

Dehaene, S. (2001). Precis of the number sense. *Mind and Language, 16*, 16–36.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources, of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284*, 970–974.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1–38.

Dougherty, C. (2003). Numeracy, literacy and earnings: Evidence from the National Longitudinal Survey of Youth. *Economics of Education Review, 22*, 511–521.

Dowker, A. (1997). Young Children's Addition estimates. *Mathematical Cognition, 3*, 141–154.

Eccles, J., Janis, E., & Harold, R. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues, 46*, 183–201.

Foorman, B. R. (2003). *Preventing and remediating reading difficulties: Bringing science to scale*. Baltimore: York Press.

Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem solving instruction among students with risk for mathematics disability with and without risk for reading disability. *Journal of Learning Disabilities, 37*, 293–306.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*, 493–513.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology, 49*, 363–383.

Geary, D. C. (1994). *Children's mathematical development: Research and practical applications*. Washington, DC: American Psychological Association.

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist, 50*, 24–37.

Geary, D. C. (1998). *Male, female: The evolution of human sex differences*. Washington, DC: American Psychological Association.

Geary, D. C. (2000). From infancy to adulthood: The development of numerical abilities. *European Child and Adolescent Psychiatry, 9*, II11–II16.

Geary, D. C. (2003). Learning disabilities in arithmetic: Problem solving differences and cognitive deficits. In H. L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 199–212). New York: Guilford Publishers.

Geary, D. C., Bow-Thomas, C., & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically-disabled children. *Journal of Experimental Child Psychology, 54*, 372–391.

Geary, D. C., Brown, S. C., & Samaranayake, V. A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed of processing differences in normal and mathematically disabled children. *Developmental Psychology, 27*, 787–797.

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology, 77*, 236–263.

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., & Onard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33*, 18–28.

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.

Ginsburg, H. P. (1989). *Children's arithmetic*. Austin, TX: PRO-ED.

Ginsburg, H. P. (1997). Mathematics learning disabilites: A view from developmental psychology. *Journal of Learning Disabilities, 30*, 20–33.

Ginsburg, H. P., & Golbeck, S. L. (2004). Thoughts on the future of research on math and science education and learning. *Early Childhood Research Quarterly, 19*, 190–200.

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* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Good, R. H., Simmons, D., Kame'enui, E., Kaminski, R. A., & Wallis, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade*. (Technical Report No. 11). Eugene, OR: University of Oregon.

Gray, E. M., & Tall, D. O. (1994). Duality, ambiguity, and flexibility: A ''proceptual'' view of simple arithmetic. *Journal for Research in Mathematics Education, 26*(2), 116–140.

Griffin, S. (2002). The development of math competence in the preschool and early school years: Cognitive foundations and instructional strategies. In J. M. Roher (Ed.), *Mathematical Cognition* (pp. 1–32). Greenwich, CT: Information Age Publishing, Inc.

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

Griffin, S. A., & Case, R. (1997). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education, 3*, 1–49.

Griffin, S. A., Case, R., & Siegler, R. S. (1994). Classroom lessons: Intergrating cognitive theory and classroom practice. In K. McGilly (Ed.), *Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure* (pp. 25–49). Cambridge, MA: MIT Press.

Hanich, L., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology, 93*(3), 615–626.

Hedeker, D. (2004). An introduction to growth modeling. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 215–234). Thousand Oaks, CA: Sage Publications.

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

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.

Jordan, N. C., Blanteno, L. M., & Uberti, H. Z. (2003). Mathematical thinking and learning disabilities. In A. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Recent research and theory* (pp. 359–383). Mahwah, NJ: Erlbaum.

Jordan, N. C., & Hanich, L. B. (2003). Characteristics of children with moderate mathematics deficiencies: A longitudinal perspective. *Learning Disabilities: Research and Practice, 18*(4), 213–221.

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

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

Jordan, N. C., Huttenlocher, J., & Levine, S. C. (1992). Differential calculation abilities in young children from middle- and low-income families. *Developmental Psychology, 28*, 644–653.

Jordan, N. C., Huttenlocher, J., & Levine, S. C. (1994). Assessing early arithmetic abilities: Effects of verbal and nonverbal response types on the calculation performance of middle- and low-income children. *Learning and Individual Differences, 6*(4), 413–432.

Jordan, N. C., Kaplan, D., & Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology, 94*, 586–597.

Jordan, N. C., Levine, S. C., & Huttenlocher, J. (1994). Development of calculation abilities in middle- and low-income children after formal instruction in school. *Journal of Applied Developmental Psychology, 15*, 223–240.

Kaplan, D. (2003). Methodological advances in the analysis of individual growth with relevance to education policy. *Peabody Journal of Education, 77*, 189–215.

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

Leppanen, U., Niemi, P., Aunola, K., & Nurmi, J. E. (2004). Development of reading skills among preschool and primary school pupils. *Reading Research Quarterly, 39*, 72–93.

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

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. Oxford, UK: Oxford University Press.

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology, 35*, 164–174.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics, 77*, 251–266.

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 Publications.

Muthén, B. O., Khoo, S. T., Francis, D., & Boscardin, C. K. (2002). Analysis of reading skills development from

kindergarten through first grade: An application of growth mixture modeling to sequential processes. In S. R. Reise & N. Dunn (Eds.), *Multilevel modeling: Methodological advances, issues, and application.* Mahway, NJ: Lawrence Erlbaum.

Muthén, L. K., & Muthén, B. (2004). *Mplus: Statistical analysis with latent variables (Version 3.0).* Los Angeles: Muthén & Muthén.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55,* 463–469.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods, 4,* 139–157.

National Assessment of Educational Progress (2004). *The nation's report card. Mathematics highlights 2003* (No. NCES2004-451) Jessup, MD: US Department of Education.

Okamoto, Y., & Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development, 61*(1–2).

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousands Oaks, CA: Sage Publications.

Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Marchant III, H. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology, 24,* 181–266.

Rubenstein, R. N. (1985). Computational estimation and related mathematical skills. *Journal for Research in Mathematics Education, 16,* 106–119.

Saxe, G. B., Guberman, S. R., Gearhart, M., Gelman, R., Massey, C. M., & Rogoff, B. (1987). Social processes in early number development. *Monographs of the Society for Research in Child Development, 52*(2).

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75,* 428–444.

Starkey, P., Klein, A., & Wakeley, P. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly, 19,* 99–120.

Teaching Integrated Mathematics and Science (TIMS) Curriculum (2004). *Math trailblazers grade K* (2nd ed.). Chicago, IL: Kendall/Hunt Publishing Company.

Threfall, J., & Frobisher, L. (1999). Patterns in processing and learning addition facts. In A. Orton (Ed.), *Pattern in the teaching and learning of mathematics* (pp. 39–46). London: Cassell.

Torgeson, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40,* 7–26.

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36,* 155–193.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 27,* 749–750.

Xu, F., & Spelke, E. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74,* B1–B11.

Appendix: Growth Mixture Modeling

This appendix provides a brief technical overview of growth mixture modeling. The growth mixture model is part of a general analytic framework for the analysis of continuous and categorical latent variables. For an overview of the general analytic framework including growth mixture modeling, see Muthén (2004).

In line with the discussion given in Kaplan (2003) growth mixture modeling begins by combining conventional growth curve modeling (e.g. Raudenbush & Bryk, 2002) with latent class analysis (e.g., Clogg, 1995) under the assumption that there exists a mixture of populations each defined by unique growth trajectories. In this respect, growth mixture modeling is more flexible than conventional growth curve modeling which assumes a single population generating the empirical growth trajectories.

An extension of latent class analysis sets the groundwork for growth mixture modeling. Specifically, latent class analysis can be applied to repeated measures at different time points. This is referred to as latent class growth analysis (LCGA) (see e.g., Nagin, 1999). As with latent class analysis, LCGA assumes homogenous growth within classes. Growth mixture modeling, by contrast, relaxes the assumption of homogenous growth within classes and is capable of capturing two important forms of heterogeneity. The first form of heterogeneity is captured by individual differences in growth through the specification of the conventional growth curve model. The second form of heterogeneity is more fundamental—representing heterogeneity in classes of growth trajectories.

The specification of the growth mixture model is similar to that given for the conventional growth curve model. The difference, as noted, lies in allowing there to be different growth trajectories for different classes. Specifically, we hypothesize there exists a latent categorical variable c_i composed of K classes ($k = 1, 2, \ldots K$) The ith student belongs to one of the K classes and the classes are mutually exclusive and exhaustive. The growth parameters are obtained via the conventional growth curve model defined as

$$y_{ti} = \eta_{0i} + \eta_{1i}a_{ti} + e_{ti}, \qquad (1)$$

where y_{ti} is a continuous outcome measure at time t for student i, a_{ti} is a time measure for person i such as age, grade, time point of study, etc., where for simplicity, we will require $a_{ti} = a_t$ meaning that the time metric is common to all individuals. The random growth factors η_{0i} and η_{1i} represent the status and rate of growth, respectively, and e_{ti} is a random error component. An acceleration parameter can also be accommodated in this specification.

The specification of the growth mixture model allows the random growth factors η_{0i} and η_{1i} to vary over latent classes through their means, variances, and covariances. Specifically, let

$$\eta_{0i} = \alpha_{0k} + \gamma'_{0k}x_i + \zeta_{0i},$$
$$\eta_{1i} = \alpha_{1k} + \gamma'_{1k}x_i + \zeta_{1i},$$

where α_{0k} and α_{1k} are the average intercept and average rate of growth that vary over the classes, respectively. The parameters γ_{0k} and γ_{1k} relate the time-invariant covariates x to the random growth factors and also allowed to vary over classes, and the ζ's are random disturbance terms which are assumed to be constant over the latent classes. The specification in Equation (2) demonstrates the flexibility of the growth mixture modeling approach—namely that the role of covariates in predicting growth may be different across substantively different trajectory classes.

The growth mixture model also allows covariate information contained in x to influence class membership directly via a multinomial logistic regression. Specifically, the multinomial logistic regression giving the probability of membership in class k given background covariate information x can be written as

$$p(c_i = k|x_i) = \frac{e^{\beta_{0k}+\beta_{1k}x_i}}{\sum_{k=1}^{K} e^{\beta_{0k}+\beta_{1k}kx_i}}, \qquad (3)$$

where a given class K is chosen as the reference class, with coefficients β_{ok} and β_{1k} fixed to zero.

Estimation of the growth mixture model is carried out using maximum likelihood of the EM algorithm (Dempster, Laird, & Rubin, 1977) as implemented in the Mplus software program (Muthén & Muthén, 2004). Technical details can be found in Muthén and Shedden (1999).

A sequence of modeling steps have been suggested for the application of growth mixture modeling and is utilized in this paper. To begin, the researcher specifies an a priori number of trajectory classes based on theory or inspection of empirical growth trajectories. 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 are interested in the smallest BIC among a variety of competing specifications. A plot of the BIC values against models that specify varying numbers of classes can be used to ascertain the number of classes. The model with the lowest BIC value is recommended.

Second, one can look at the posterior probability of being assigned to a particular class. The precision of group assignment is suggested by classes with substantively large numbers of assigned units.

Third, 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.

Following the determination of the number of latent classes, one can add predictors of the latent classes model, thus fully specifying a growth mixture model. However, the addition of covariate information can influence latent trajectory class determination, and so it is essential to re-examine the statistical and substantive utility of the latent trajectory classes using the methods just outlined.