Co-development of fraction magnitude knowledge and mathematics achievement from fourth through sixth grade

Article *in* Learning and Individual Differences · December 2017

DOI: 10.1016/j.lindif.2017.10.005

Luke F Rinne

Nancy C Jordan
University of Delaware

Ai Ye
University of North Carolina at Chapel Hill

Ilyse Resnick
Pennsylvania State University Lehigh Valley

Learning and Individual Differences 60 (2017) 18–32

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Co-development of fraction magnitude knowledge and mathematics achievement from fourth through sixth grade[☆]

Nicole Hansen[a], Luke Rinne[b], Nancy C. Jordan[b,*], Ai Ye[b], Ilyse Resnick[c], Jessica Rodrigues[b]

[a] Fairleigh Dickinson University, Teaneck, NJ, United States
[b] University of Delaware, Newark, DE, United States
[c] Penn State University, Lehigh Valley, Center Valley, PA, United States

ARTICLE INFO

Keywords:
Fractions
Mathematics achievement
Numerical magnitude
Estimation
Comparison

ABSTRACT

Fraction magnitude understanding is linked to student achievement in mathematics, but the direction of the relation is not clear. To assess whether fraction magnitude knowledge and mathematics achievement develop in a bidirectional fashion, participants ($N = 536$) completed a standardized mathematics achievement test and two measures of fraction magnitude understanding—fraction comparisons and fraction number line estimation (FNLE)—twice yearly in 4th–6th grades. Cross-lagged panel models revealed significant autoregressive paths for both achievement and magnitude knowledge, indicating longitudinal stability after accounting for correlational and cross-lagged associations. Mathematics achievement consistently predicted later FNLE and fraction comparison performance. FNLE and fraction comparisons predicted mathematics achievement at all time points, although this relation diminished over time. Findings suggest that fraction magnitude knowledge and broader mathematics achievement mutually support one another. FNLE predicted subsequent mathematics achievement more strongly than did fraction comparisons, possibly because the FNLE task is a more specific measure of fraction magnitude understanding.

1. Introduction

A large body of research links student understanding of fraction magnitudes to broader mathematics achievement (Bailey, Hoard, Nugent, & Geary, 2012; Resnick et al., 2016; Siegler, Thompson, & Schneider, 2011). Fraction magnitude understanding involves the ability to comprehend, estimate, and compare the sizes of fractions (Fazio, Bailey, Thompson, & Siegler, 2014). However, the direction of the relation between mathematics achievement and fraction magnitude understanding is not clear.

Some research indicates that knowledge of fraction magnitudes provide a key underlying structure for later learning of related mathematics skills (Booth & Newton, 2012). For example, knowledge of fraction equivalence (e.g., $1/3 = 4/12$) supports learning about ratios and proportions (Siegler & Pyke, 2013). Further, understanding how the numerator and denominator of a fraction work together to determine its magnitude supports algebra learning for problems such as finding the slope of a line, measuring rates, and manipulating algebraic equations (Booth & Newton, 2012). However, it is also plausible that fraction magnitude understanding is in large part a function of prior mathematics achievement (Bailey, Siegler, & Geary, 2014; Vukovic et al., 2014). Whole number magnitude understanding, calculation fluency, multiplicative reasoning, and long division all predict later development of fraction knowledge, including fraction magnitude understanding (Hansen et al., 2015; Jordan et al., 2013; Vukovic et al., 2014).

The possibility of a bidirectional relation between fraction magnitude understanding and broad mathematics achievement—that is, a relationship in which learning in one area supports learning of the other, and vice versa—has not been fully investigated. Modest indirect evidence of such a bidirectional relation comes from Watts et al. (2015), who examined the effect of first-grade mathematics achievement on later mathematics achievement at age 15. Results showed that fraction knowledge in fifth grade mediated this relationship, suggesting that early mathematics achievement facilitates fraction learning, which in turn bolsters mathematics achievement more broadly. However, data in the form of repeated, concurrent measures of fraction knowledge and mathematics achievement is necessary to more rigorously test for a longitudinal, mutually supportive relationship.

☆ This research was supported by Awards R324C100004 from the U.S. Institute of Education Sciences.
* Corresponding author at: School of Education, University of Delaware, 19716, United States.
E-mail address: njordan@udel.edu (N.C. Jordan).

http://dx.doi.org/10.1016/j.lindif.2017.10.005
Received 18 October 2016; Received in revised form 5 September 2017; Accepted 7 October 2017
1041-6080/

1.1. Cross-lagged panel models of relationships between numerical knowledge and mathematics achievement

Cross-lagged panel models have been used to examine the directionality of the relation between whole number knowledge and broad mathematics achievement from kindergarten to second grade (Friso-van den Bos et al., 2015), as well as the relation between fraction knowledge and broad mathematics achievement in sixth and seventh grade (Bailey et al., 2012). Cross-lagged panel modeling provides better evidence of directionality than other correlational and mediation designs, as this approach allows for the evaluation of cross-domain lagged effects while controlling for effects of prior achievement within each domain (i.e., autoregressive effects). Thus, cross-lagged models can be used to analyze phenomena longitudinally and investigate potential causality in both directions (Bollen & Curran, 2006).

Friso-van den Bos et al. (2015) used cross-lagged panel analyses to examine the relation between whole number magnitude understanding, measured by a whole number line estimation (WNLE) task (using a 0–100 number line) and mathematics achievement (measured by a standardized achievement test covering a wide range of mathematics topics) over four time points in first and second grades. It was shown that WNLE ability and general mathematics achievement develop in tandem over the course of first and second grades. For example, as children solve general arithmetic problems, they may use a mental number line (e.g., Halberda, Mazzocco, & Feigenson, 2008; Siegler & Booth, 2004), reasoning, for example, that 21 minus 13 is unlikely to be 34, because 34 is to the right of 21 on the number line. This, in turn, improves their understanding of numerical magnitudes (Friso-van den Bos et al., 2015).

Bailey et al. (2012) employed a cross-lagged panel model to investigate the relation between fraction magnitude knowledge (measured by a comparison task that prompted students to choose the larger of two fractions) and broad mathematics achievement in sixth and seventh grades. Fraction magnitude knowledge predicted one-year gains in mathematics achievement, but the opposite predictive relationship was not significant. In other words, prior fraction magnitude knowledge appeared to underlie gains in mathematics achievement between sixth and seventh grade, but prior mathematics achievement did not influence development of fraction magnitude skills during the same time period. Bailey and colleagues speculate that the unidirectional relationship reflects a causal mechanism whereby the development of fraction magnitude knowledge may help students learn decimals, ratios and proportions, and pre-algebraic reasoning, all key topics that form the foundation for learning algebra. However, the generalizability and validity of the findings are limited by the study's assessment of only two time points, both of which occurred *after* the bulk of fractions instruction took place in fourth through sixth grade (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

We hypothesize that, had Bailey et al. (2012) assessed directional relationships earlier in development, evidence of a bi-directional relationship might have emerged. Previous research (e.g., Vukovic et al., 2014) indicates that a myriad of whole number skills predicts later fraction knowledge (including fraction magnitude understanding). Additionally, research in the earlier grades (e.g., Friso-van den Bos et al., 2015) indicates that there is a bi-directional relation between whole number magnitude knowledge and mathematics achievement. As such, magnitude knowledge seems to develop iteratively with mathematics achievement in elementary school, in contrast to what was found by Bailey et al. (2012) with older students. One issue may be that Bailey et al. used only a brief fraction comparison task to assess fraction magnitude knowledge; this task may reflect knowledge of procedures instead of or in addition to magnitude understanding. Thus, it is possible a bi-directional relationship might be observed for a different measure of magnitude knowledge, such as fraction number line estimation (FNLE).

Previous research on the relation between fraction magnitude understanding and mathematics achievement has been limited both in quantity and scope; studies have focused on older students (e.g., Bailey et al., 2014), relied on cross-sectional designs (e.g., Siegler & Pyke, 2013), or examined whether fraction understanding predicts much later mathematics outcomes in high school (e.g., Siegler et al., 2012). To date, no longitudinal research has used multiple measures to thoroughly examine the potential for a bi-directional relation between fraction magnitude knowledge and mathematics achievement during the important period when primary instruction in fractions takes place in school.

2. The present study

Our study aims to test our hypothesis that the reciprocal relationship witnessed between whole number magnitude understanding and mathematics achievement in early schooling (Friso-van den Bos et al., 2015) also holds for fraction magnitude knowledge and mathematics achievement in the intermediate grades. First, we hypothesize that, as was found by Bailey et al. (2012), fraction magnitude understanding will predict broader math achievement. Fraction knowledge is likely to support acquisition of other more complex mathematics topics in fifth grade, such as knowledge of long division and decimals. However, in contrast to the findings of Bailey et al., we also expect that general mathematics ability (e.g., whole number knowledge) will support fraction magnitude understanding, as proficiency in other areas of mathematics seems likely to help students learn fractions. For example, in fourth grade, students learn about multiplicative relationships (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), which would then help them reason about relationships among fraction magnitudes (e.g., one-half is the same as two-fourths, and both numbers are represented in the same location on the number line). The notion of a bidirectional relationship between fraction magnitude knowledge and general mathematics achievement is consistent with Siegler's integrated theory of numerical development (Siegler et al., 2011; Siegler & Lortie-Forgues, 2014), which holds that numerical learning is a progressive broadening of the set of numbers whose magnitudes can be accurately represented. Thus, finding a mutually predictive relationship would provide further evidence for the continuous role of numerical magnitude knowledge in learning increasingly complex mathematics, such as algebra (e.g., reasoning about linear equations (Booth, Newton, & Twiss-Garrity, 2014)).

We assessed students on fraction magnitudes and mathematics achievement at two time-points per grade, resulting in a total of six measurements. Given that no prior research has focused on the relationship between fraction magnitude understanding and mathematics achievement between fourth and sixth grade, despite this being the period during which most fraction instruction takes place, the present study fills an important gap in the literature. Our study also differs from previous work by controlling for a broader range of general cognitive abilities that may influence fraction knowledge and general mathematics achievement. Previous research has either not controlled for cognitive factors at all (Friso-van den Bos et al., 2015) or has included only a few variables (Bailey et al., 2012), even though recent analyses suggest that cognitive factors (e.g., intelligence and working memory) significantly predict the development of children's broad mathematics achievement over time (Bailey, Watts, Littlefield, & Geary, 2014). Failure to control for these variables could potentially confound results regarding the relationship between fractions ability and math achievement. Selection of the control variables used in the present study was guided by prior research on predictors and correlates of mathematical achievement (Geary, 2004; Gunderson, Ramirez, Beilock, & Levine, 2012; Swanson, 2011) and fraction knowledge in particular (Hecht, Close, & Santisi, 2003; Hecht & Vagi, 2010; Jordan et al., 2013; Seethaler, Fuchs, Star, & Bryant, 2011). We control for background variables, including age, gender, and income status, as well as vocabulary, matrix reasoning, reading, working memory, and attentive behavior.

A secondary yet important aim of the present study is to compare and contrast two commonly used measures of fraction magnitude knowledge to pinpoint how each relates to broader mathematics skills: fraction number line estimation (FNLE), which requires students to estimate the locations of fraction magnitudes on a number line, and fraction comparison, which requires students to indicate which of two fractions is larger. In previous work, these two tasks have largely been treated as interchangeable measures of the construct of fraction magnitude understanding (e.g., Bailey et al., 2012; Bailey et al., 2015; DeWolf & Vosniadou, 2015; Fazio, DeWolf, & Siegler, 2016; Resnick et al., 2016; Siegler et al., 2011; Siegler & Pyke, 2013; Torbeyns, Schneider, Xin, & Siegler, 2015). However, while the FNLE task encourages students to apply their knowledge of magnitudes to place fractions accurately on a number line (Siegler et al., 2011), items on the fraction comparison task can sometimes be completed accurately with memorized procedural strategies that do not involve understanding of magnitudes (Rinne, Ye, & Jordan, 2017; Fazio et al., 2016). For example, students can identify the larger fraction simply by comparing the numerator when the denominators are the same; when two fractions have the same numerator (e.g., 4/10 vs. 4/12), students can compare denominators. Answers using these strategies, even when correct, do not necessarily indicate strong numerical magnitude understanding, as students may only be looking at one part of a fraction (e.g., numerator or denominator) rather than considering how both parts work together to represent a specific location on the number line. Older students might reach a correct answer by cross-multiplying, a taught procedure. We do not disregard the possibility that some students might use strategies to place fractions on the number line, such as numerical transformation strategies (rounding, simplifying, or converting fractions to other forms, such as decimals) or line segmentation strategies. However, these strategies still require reasoning about the magnitude of the fraction to be executed accurately. Since fraction number line estimation is not taught in schools to the same extent that fraction comparison is, we hypothesize students would be more likely to use memorized strategies when solving fraction comparisons. As such, we suspect that FNLE has higher validity as a measure of magnitude understanding per se. Specifically, we expect that there will be both shared and unshared variance in performance across the two tasks; both the FNLE and fraction comparison tasks assess fraction magnitude, but the measures will not be completely aligned. Factor analysis was used to investigate shared versus unshared variance in performance on the fraction magnitude measures.

3. Method

3.1. Participants

Data were collected as part of a larger study aimed toward understanding mathematical development (Jordan et al., 2013; Jordan, Resnick, Rodrigues, Hansen, & Dyson, 2016). The study was approved by the University's Institutional Review Board (IRB). Third-graders (n = 746) in nine public schools from two adjacent districts were sent an informed consent letter requesting their participation in the study, and all participating students and their parents provided assent.

The participating public schools served families of diverse socioeconomic backgrounds. A total of 517 students (69%) agreed to participate in this study; however, 36 switched to a non-participating school or declined to participate prior to the first assessment. Students were followed through spring of sixth grade. In sixth grade, the final year of the study, students transitioned from elementary school to middle school. The majority of participating sixth-graders (84%) attended one of four middle schools in the participating school districts; the other children attended a different middle school in or close to the original school districts. By the end of third grade, 23 students dropped out of the study, by the end of fourth grade an additional 68 students dropped out, and by the end of sixth grade an additional 39 children dropped out. Attrition was due to students moving to another school district out of the study (67%), a lack of information on students' elementary to middle school transition (23%), and students withdrawing from the study (10%). The sample was replenished in fourth grade (n = 27 new children) and again in fifth grade (n = 28 new children). In total, the sample for the present study included 536 students.

The final study sample was 47.0% male, 51.9% white, 40.0% black, 5.7% Asian/Pacific Island, and 2.5% American Indian/Alaskan Native. A separate 17.7% of children identified their ethnicity as Hispanic, according to school district records. Over half of the students (60.9%) qualified for participation in a school free or reduced lunch program and thus were considered to be low income. The mean age at the start of third grade was 8.82 (SD = 0.45) years. In the sample, 10.6% of students were identified as English learners (ELs) and an additional 10.6% of students received special education services. All participating schools followed curriculum benchmarks aligned with the Common Core State Standards in Mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) starting in fourth grade.

3.2. Control measures (assessed in third grade)

3.2.1. Vocabulary

The *Peabody Picture Vocabulary Test* (PPVT; Dunn & Dunn, 2007) was used to assess students' receptive vocabulary. The assessor said a word and asked the student to point to one of four picture choices that corresponds to the meaning of the word. The total score on this measure was the number of words correctly identified. The PPVT has high internal stability ($\alpha > 0.96$) (Cronbach, 1951) and the correlation between the PPVT and overall verbal IQ was 0.89, indicating that it is a valid measure of general verbal ability (Dunn & Dunn, 2007).

3.2.2. Matrix reasoning

The Matrix Reasoning subtest of the *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999) assessed matrix reasoning. Students were shown a series of two-by-two grids, with geometric patterns pictured in three out of the four cells. Students were asked to complete the pattern by pointing to one of five response options presented below the grid. Internal reliability of this measure is high ($\alpha > 0.90$) and the correlation between matrix reasoning and overall performance/nonverbal IQ is 0.87 (Wechsler, 1999).

3.2.3. Attentive behavior

To measure attentive behavior, students' mathematics teachers completed the inattention subscale of the *SWAN Rating Scale* (Swanson et al., 2006). The scale contained nine items based on the criteria for attention deficit hyperactivity disorder for inattention from the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition* (American Psychiatric Association, 1994). For each item, teachers rated children's attention during their mathematics classes on a scale of 1 (below average) to 7 (above average). The inattention subscale had a high internal consistency in this sample ($\alpha > 0.97$) and was correlated with teachers' subsequent ratings at fourth, fifth, and sixth grades (r ranges from 0.52 to 0.67). Attentive behavior is highly predictive of mathematics outcomes (Fuchs et al., 2006, 2010; Vukovic et al., 2014).

3.2.4. Working memory

Working memory was assessed with the Counting Recall subtest of the *Working Memory Test Battery for Children* (WMTB-C; Pickering & Gathercole, 2001). Students were asked to recall the number of dots in a series of arrays in the order of presentation. For example, if presented with four dots, two dots, and then six dots, the correct response would be “four, two, six”. After successfully completing three out of six trials, the number of arrays increased by one.

The test-retest reliability for this task is 0.61 (Pickering & Gathercole, 2001), and this task is predictive of mathematics outcomes (Geary, 2011).

3.2.5. Reading

In the Sight Word Efficiency subtest of the *Test of Word Reading Efficiency* (TOWRE; Torgesen, Wagner, & Rashotte, 1999), students were presented with a list of 104 written words, and given 45 s to read aloud as many words as possible from the list. The score was the total number of correctly read words. In third grade, the test-retest reliability for this task is 0.97.

3.2.6. Demographic variables

Demographic variables included age at third grade entry, income status (i.e., whether children qualified to participate in the free/reduced lunch program at school), and gender.

3.3. Outcome measures (assessed in fourth through sixth grades)

3.3.1. Fraction number line estimation

On the fraction number line estimation task (FNLE; adapted from Siegler et al., 2011 and Siegler & Pyke, 2013), students estimated the location of fractions on 0–1 and 0–2 number lines. Each number line was presented in the middle of the screen on a laptop computer using Direct RT v2012 and was 17.5 cm long. Fractions were presented one at a time in a fixed order below the middle of the number line. For each item, the cursor was set at "0", and students used the arrow keys to slide the cursor along the number line. Students pressed a key to indicate their response. After each estimate, a new blank number line and a new fraction were presented and the cursor was reset to "0".

For the 0–1 number line task, students began by observing the assessor demonstrate where 1/8 is located. Students then completed a practice trial estimating the fraction 1/4 without feedback. Next, students estimated the locations of 1/5, 13/14, 2/13, 3/7, 5/8, 1/3, 1/2, 1/19, and 5/6. The same procedure was used on the 0–2 number line task. The assessor modeled where 1/8 and 1 1/8 are located. Students then estimated a practice fraction (1/4). For the 0 to 2 number line, the students estimated the location of the fractions and mixed numbers 1/3, 7/4, 12/13, 1 11/12, 3/2, 5/6, 5/5, 1/2, 7/6, 1 2/4, 1, 3/8, 1 5/8, 2/3, 1 1/5, 7/9, 1/19, 1 5/6, and 4/3. In total, students estimated the locations of 28 fractions and mixed numbers on FNLE task.

To score the measure, the percent absolute error (PAE) was calculated by dividing the absolute difference between the estimated and actual magnitudes by the numerical range of the number line (1 or 2), and then multiplying by one hundred for each estimate (Siegler et al., 2011). Each student was assigned a single score that represented his or her mean PAE on the task. Scores were then multiplied by -1 so that higher scores indicated higher achievement. We opted to use PAE rather than linear R-square for ease of interpretation and also because some children's estimates can fit a linear function despite being highly inaccurate (R. Siegler, personal communication, May 20, 2016).

Coefficient alpha on the 0–1 and 0–2 number line task for our sample was 0.870 (winter of fourth grade), 0.914 (spring of fourth grade), 0.980 (fall of fifth grade), 0.979 (spring of fifth grade), and 0.975 (winter of sixth grade).

3.3.2. Fraction comparisons

Fraction comparisons (Comp) were assessed with an expanded and adapted version of the timed paper and pencil task used by Bailey et al. (2012). Students circled the larger of two fractions on 24 items within 3 min. Six types of fraction comparisons (four items of each type) were included in the measure: (a) unit fractions (1/3 vs. 1/2, 1/55 vs. 1/57, 1/4 vs. 1/5, 1/10 vs. 1/100), (b) fractions with like denominators (7/12 vs. 9/12, 5/7 vs. 6/7, 24/48 vs. 28/48, 2/10 vs. 4/10), (c) fractions with like numerators (5/3 vs. 5/2, 2/4 vs. 2/5, 6/9 vs. 6/12, 3/7 vs. 3/8), (d) comparison to one fraction equivalent to one-half (50/100 vs. 16/17, 20/40 vs. 8/9, 5/10 vs. 3/4, 10/20 vs. 5/6), (e) reciprocal fractions (3/2 vs. 2/3, 8/4 vs. 4/8, 5/14 vs. 14/5, 5/6 vs. 6/5), and (f) fractions with similar denominators (3/10 vs. 2/12, 12/50 vs. 8/60, 6/33 vs. 9/30, 6/8 vs. 3/9). The internal reliabilities of the fractions comparison measure for our sample for the six time points in the present study were 0.867 (fall of fourth grade), 0.896 (spring of fourth grade), 0.912 (fall of fifth grade), 0.922 (spring of fifth grade), 0.921 (fall of sixth grade), and 0.919 (spring of sixth grade).

3.3.3. Mathematics achievement

To measure broad mathematics achievement, we obtained children's scores from the mathematics portion of the Delaware Comprehensive Assessment System (DCAS; American Institutes of Research, 2012). The mathematics section of the DCAS required students to answer grade-level multiple-choice questions that are aligned with the Common Core State Standards. The DCAS was a broad mathematics achievement measure that assessed knowledge in several domains, including algebraic reasoning (e.g., find a given term in an arithmetic sequence), numeric reasoning (e.g., use and apply meanings of multiplication and division; develop an understanding of fractions as parts of unit wholes and division of whole numbers), geometric reasoning (e.g., analyze and classify two-dimensional shapes according to their properties), and quantitative reasoning (e.g., construct and use data displays) (Delaware Department of Education, 2013). We used the accountability score at each grade level, which is computed using grade-level items. Scores range from 0 to 1300. Published internal consistency at each time point was 0.87 (fall of fourth grade), 0.86 (spring of fourth grade), 0.85 (fall of fifth grade), 0.89 (spring of fifth grade), 0.88 (fall of sixth grade), and 0.88 (spring of sixth grade; American Institutes of Research, 2012).

3.4. Procedure

Demographic data were obtained from the participating school districts. Control measures of vocabulary, matrix reasoning, attentive behavior, working memory, and reading were administered in third grade. The fraction comparisons measure was administered at six time points in the fall and spring of each school year (fourth, fifth, and sixth grade). The fraction number line estimation measure was given at five time points: winter and spring of fourth grade, fall and spring of fifth grade, and winter of sixth grade. The mathematics achievement measure was given during the fall and spring of each school year (fourth, fifth, and sixth grade).

All measures were administered by trained assessors on our research team, with the exception of the mathematics achievement test, which was administered by the school district according to published guidelines (American Institutes of Research, 2012).

3.5. Data analytic procedure and preliminary analyses

We conducted a preliminary confirmatory factor analysis (CFA) to determine whether the fraction comparisons and FNLE measures tap into a shared construct. We evaluated the fit of the model using the chi-square (χ^2) statistic, the Root Mean Square Error of Approximation (RMSEA; < 0.08 adequate, < 0.05 excellent; Brown & Cudeck, 1993); the Comparative Fit Index (CFI; > 0.90 adequate, > 0.95 excellent; Bentler, 1990; Bentler & Bonett, 1980), and the Tucker Lewis Index (TLI; > 0.90 adequate, > 0.95 excellent; Tucker & Lewis, 1973). A CFA of five factors (one at each time point) regressed on the corresponding total scores (i.e., FNLE, fraction comparisons), respectively, indicated an adequate fit: χ^2 (45) = 4395.127, RMSEA = 0.128, CFI = 0.949, TLI = 0.909. (A factor score regressing on two indicators is sufficient according to Little, Lindenberger, & Nesselroade, 1999.) This provides strong evidence that both the FNLE and comparisons measures are tapping into a single construct corresponding to fraction magnitude understanding. However, a closer look at factor loadings revealed that the factor at every time point was dominated by FNLE: the factor loadings for FNLE were appreciably higher than those of fraction comparisons. That is, the variances captured by the factor at each time point ranged from 72% to 83% for FNLE, but only 30% to 57% for fraction comparisons. This indicates that a magnitude understanding construct explains considerable variance in performance on both measures, but there still remains a small portion of FNLE performance and a larger portion of fractions comparisons performance unexplained. It is possible that the additional unexplained variance in the comparisons task could contribute to differences in results for the cross-lagged panel model. As such, we proceeded to analyze the relationships between each fraction measure and mathematics achievement separately. Similar results for models with each measure would suggest that the observed relation between mathematics achievement and fraction magnitude knowledge is broadly valid, regardless of whether fraction magnitude understanding is measured via comparisons or FNLE. However, divergent results would suggest that different aspects of fraction magnitude understanding and broad mathematics achievement relate to one another in different ways, presenting a more complex scenario.

An autoregressive, cross-lagged panel model (AR-CL) design allowed us to deliberately consider potential bidirectional relationships between two constructs over multiple time points (Bollen & Curran, 2006; de Jonge et al., 2001; Farrell, 1994; Finkel, 1995; Zapf, Dormann, & Frese, 1996). This type of model allows for simultaneous examination of longitudinal influences of one construct on another construct and vice versa (i.e., lagged effects), while also controlling for concurrent correlations between constructs along with the stability of each construct over time (i.e., autoregressive effects, or the influence of prior scores for a given measure; Bollen & Curran, 2006; Selig & Little, 2012). Although the AR-CL model cannot conclusively demonstrate causality, it does permit us to explore and compare plausible causal hypotheses (Bollen & Curran, 2006; Burkholder & Harlow, 2003; Farrell, 1994). That is, the AR-CL model can specify the effects of a given type of knowledge (i.e., mathematics achievement or fraction magnitude understanding) on future measurements of the other type of knowledge, while controlling for knowledge of both types at a previous time points. We can also control for other variables in the model that may be related to the types of knowledge being studied; in the present case, we include background variables.

Although cross-lagged models have been critiqued elsewhere (see Berry & Willoughby, 2016), we opted to use them in the present study because they fit our data well and allowed us to address our research questions most clearly and concisely. Although more complex models might also be useful, a primary aim of our study was to directly compare our results with previous work that has used cross-lagged panel models. The main focus of the present study was on the directionality of predictions (between fraction magnitude understanding and mathematics achievement), rather than growth curves or explanations of state versus trait variance (for further discussion of state versus trait variance in mathematics achievement, see Bailey et al., 2014). Specifically, we were interested in the strength of relationships between fraction understanding and achievement across measures, as well as changes over time, evidenced by lagged effects that persist above and beyond autoregressive effects and concurrent correlations. Cross-lagged panel models are adequate for addressing these particular questions.

We constructed two sets of AR-CL models in alignment with the results of the preliminary confirmatory factor analysis. The first set of AR-CL models investigated the directionality of relationships between fraction number line estimation (FNLE) accuracy and mathematics achievement (Ach) while controlling for background variables as well as prior knowledge. The second set of AR-CL models investigated the directionality of relationships between fraction comparisons (Comp) accuracy and mathematics achievement while controlling for background variables as well as prior knowledge. Following the recommendation of Zapf et al. (1996) and de Jonge et al. (2001), we conducted a systematic evaluation of a set of nested intermediate models for each measure of fraction understanding as outlined below and shown in Fig. 1:

3.5.1. Model 1

An independent, first-order autoregressive model. This model contained the effects of background variables on initial scores, contemporaneous correlations between each measure of fraction knowledge (FNLE or Comp) and mathematics achievement, and first-order autoregressive paths among longitudinal measurements of fraction knowledge and mathematics achievement. That is, in this model, each longitudinal test score was regressed only on the immediately preceding score on the same test, with no crossing paths connecting scores on different measures.

3.5.2. Model 2

An independent, higher-order autoregressive model. This model contained the relationships described above, as well as higher-order paths from each mathematics achievement score to *all* prior mathematics achievement scores (rather than just the immediately preceding score). These additional autoregressive paths were included to account for the fact that the mathematics achievement test differed from year to year (reflecting differences in content taught). It is possible that early acquisition of certain forms of mathematics content has downstream effects on later test scores that are not captured by scores at immediately subsequent time points. Higher-order autoregressive paths were not included for measures of fraction knowledge (FNLE and Comp) because these tests were identical at each time point and always assessed the same content, making higher-order autoregressive relationships less plausible.

3.5.3. Model 3a

(Ach → FNLE/Comp) A higher-order autoregressive model (as in Model 2) with unidirectional paths from each measurement of mathematics achievement to the measurement of each type of fraction knowledge at the subsequent time point. This model assesses fraction knowledge as a function of mathematics achievement.

3.5.4. Model 3b

(FNLE/Comp → Ach) A higher-order autoregressive model (as in Model 2) with unidirectional paths from each measurement of fraction knowledge to the measurement of mathematics achievement at the subsequent time point. This model assesses mathematics achievement as a function of fraction knowledge.

3.5.5. Model 4

(FNLE/Comp ↔ Ach) A higher-order autoregressive model (as in Model 2) with cross-lagged paths from each measurement of fraction knowledge (FNLE/Comp) to the measurement of mathematics achievement at the subsequent time point, and vice versa. This model assesses the bidirectional relationship between mathematics achievement and fraction knowledge.

We used the statistical package Stata to construct models for both FNLE and Comp corresponding to Models 1–4 above. These models were fit using maximum likelihood estimation with missing values (i.e., full-information maximum likelihood estimation), which uses all the data that are available in order to estimate the model without imputation (StataCorp, 2013). Standard errors were based on the observed information matrix (OIM). We evaluated the fit of each model using the chi-square (χ^2) statistic, the Comparative Fit Index (CFI; > 0.90 adequate, > 0.95 excellent; Bentler, 1990; Bentler & Bonett, 1980), the Tucker Lewis Index (TLI; > 0.90 adequate, > 0.95 excellent; Tucker & Lewis, 1973), and the Root Mean Square Error of Approximation (RMSEA; < 0.08 adequate, < 0.05 excellent; Brown & Cudeck, 1993). We also utilized the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and likelihood ratio (LR) tests to compare model fit at each stage to that of previous nested models;

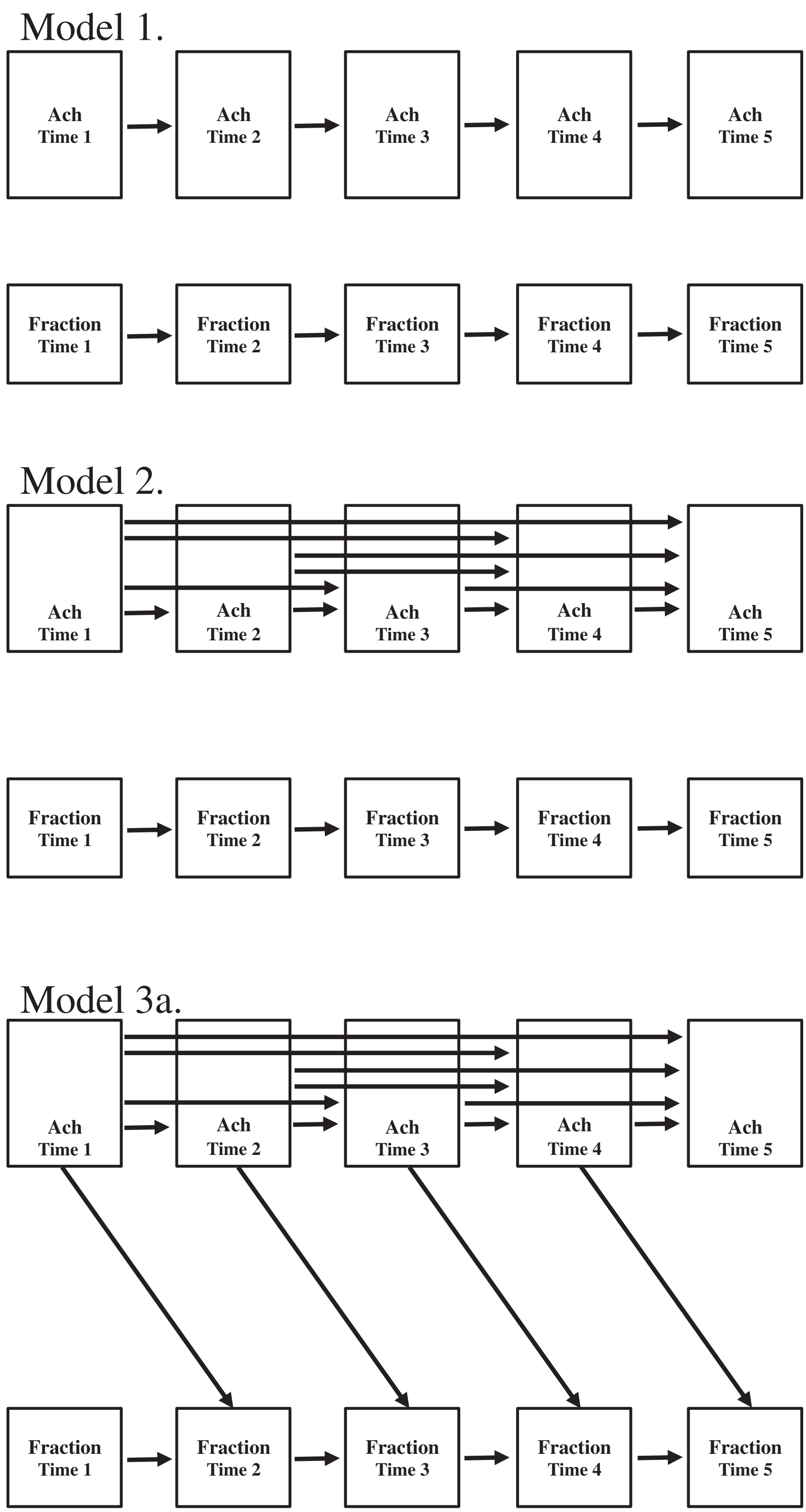
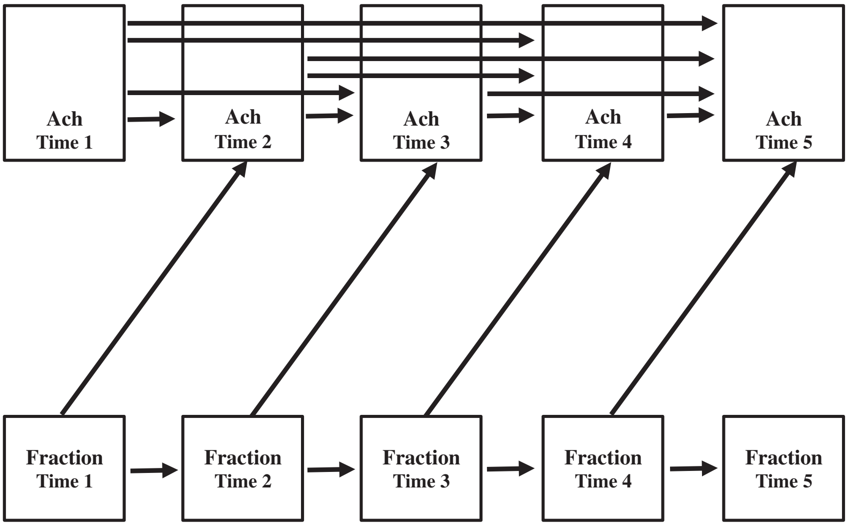


Fig. 1. Procedure for developing final autoregressive, cross-lagged models. Not pictured: effects of covariates and contemporaneous correlations. Ach = Mathematics achievement. Fraction = Fraction magnitude understanding, as measured either by fraction number line estimation or fraction comparison.

lower AIC and BIC values indicate better model fit. The likelihood ratio test indicates whether a model produces a significantly better fit than a simpler nested model. Because model selection aims to balance parsimony with fit (Kline, 1998), paths should only be added to the model if they significantly improve model fit.

After determining which model structure (1–4) provided the best fit to the data, we then planned to examine individual standardized path weights to compare results for the FNLE and comparisons models, as well as to observe potential changes over time within each model. However, Models 1–4 do not account for potential clustering of scores among students who share the same teacher in a given year. With such nested data, between-group differences in the outcome variable may induce within-group correlation, which may bias standard errors and thus the test statistics Therefore, we subsequently used M*plus* 7.1 to test

Model 3b.



Model 4.

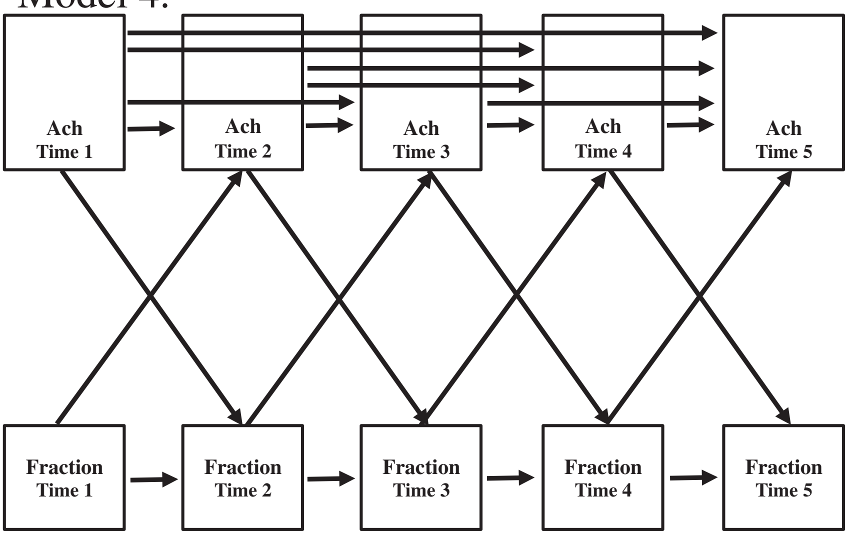


Fig. 1. (*continued*)

whether random slopes for effects of teacher on each dependent measure at each time point exhibited significant variance. If a random slope exhibited significant variance for a given measure/time point, which indicates a cluster effect, we then additionally regressed this random slope on the full set of covariates to investigate whether covariate values may exaggerate or mitigate the effect of teacher on performance for a given measure. If random slopes for teacher effects did not exhibit significant variance (suggesting a null cluster effect), they were not included in the final model (Model 5). We adopted this approach with the purpose to account for potential cluster effects; again, evaluation of within- versus between-group effects either for this particular sample or for a larger population is not part of the scope of the present study.

4. Results

Means and standard deviations for all measures are provided in Table 1. Correlations among all dependent variables are shown in Table 2. All correlations were statistically significant. Generally, correlations between the same measure at different time points were strong, while correlations between each of the two fraction measures and mathematics achievement were moderate-to-strong throughout the study.

To investigate whether dropping out of the longitudinal study was systematically related to any of the covariates in the model, we tested for covariate-dependent missingness using the procedure described by Little (1995). Results were not significant for any of the dependent variables: Achievement, $\chi^2(423) = 240.269$, *ns*; Comparison, $\chi^2(675) = 352.782$, *ns*; or FNLE, $\chi^2(387) = 119.286$, *ns*. Thus, missing data patterns were unrelated to covariates in the model.

4.1. Relations between fraction number line estimation and mathematics achievement

We constructed a series of nested panel models to assess the overall structure of the relationship between FNLE and mathematics achievement. Fit indices for all intermediate models (Table 3) show that the

Table 1
Means and standard deviations on all measures.

Measure	M	SD
General measures		
Vocabulary (PPVT; percentile)	47.16	28.63
Matrix reasoning (WASI; scaled score [M = 10])	9.81	3.26
Working memory (WMTB-C; 33)	19.38	21.31
Reading (TOWRE; percentile)	64.98	21.31
Attentive behavior (SWAN, 63)	36.76	21.01
Fraction number line estimation (0–1 and 0–2; 28; percent absolute error)		
Winter fourth grade	24.45	8.51
Spring fourth grade	19.16	9.33
Fall fifth grade	19.16	10.46
Spring fifth grade	15.75	10.26
Winter sixth grade	12.96	9.52
Fraction comparisons (24)		
Fall fourth grade	11.46	5.43
Spring fourth grade	14.88	6.23
Fall fifth grade	14.77	6.57
Spring fifth grade	16.51	6.58
Winter sixth grade	17.64	6.26
Spring sixth grade	18.14	6.07
Mathematics achievement (accountability score)[a]		
Fall fourth grade	680.63	60.45
Winter fourth grade	701.30	75.39
Spring fourth grade	761.12	76.81
Fall fifth grade	715.73	68.22
Spring fifth grade	788.83	65.54
Fall sixth grade	741.94	62.16
Spring sixth grade	793.65	75.79

Note. Numbers in parentheses indicate total number of items.

[a] Mean scores at each grade level were somewhat below standard at the fall and winter time points. Each year, the mean score met standards for mathematics at the spring time point using DCAS performance cut scores for mathematics (American Institutes for Research, 2012).

best-fitting model was the full higher-order autoregressive cross-lagged model (Model 4), χ^2 (82) = 296.939, RMSEA = 0.070, CFI = 0.952, TLI = 0.928, AIC = 55,158.441, BIC = 55,616.843. A significant likelihood ratio test also indicates that including the full set of cross-lagged paths significantly improved the fit of the higher-order autoregressive model (Model 4) over that of both unidirectional models. This finding suggests that the relationship between FNLE and mathematics achievement from fourth to sixth grade is bidirectional, even after controlling for autoregressive effects, contemporaneous correlations, and effects of background characteristics.

To account for potential clustering of observations among students sharing the same teacher, we then tested whether adding random slopes for the effect of teacher exhibited significant variance. Dummy codes were used to identify teachers, and for each measure at each time point, a random slope was regressed on the set of dummy variables. This procedure was repeated for each measure at each time point, with random slopes only being retained in the model if they showed significant variance. For the FNLE/Achievement model, the only random slope with significant variance was the one associated with the FNLE measure in the spring of 4th grade, $s^2 = 16.78$, $SE = 5.16$, $p = 0.001$. The variances of random slopes for the effect of teacher were non-significant (all $ps > 0.05$) at all other time points for both FNLE and achievement. In the final model (Model 5), the slope of the teacher effect was further regressed on the full set of covariates.

Table 4 gives covariances among background variables for Model 5, as well as path coefficients that relate each background variable to baseline FNLE scores, baseline mathematics achievement scores, and the random slope of the teacher effect in the spring of fourth grade. Gender, vocabulary, matrix reasoning, working memory, and attentive behavior were significant predictors of baseline FNLE scores, while age, vocabulary, matrix reasoning, reading, working memory, and attentive behavior were significant predictors of baseline math achievement. We also observed positive effects of age and gender on the random slope of the teacher effect in spring of fourth grade, indicating that older students and females were more susceptible to effects of teacher (either positive or negative).

Fig. 2 shows a path diagram with standardized coefficients for longitudinal relationships. Mathematics achievement predicted FNLE skill at each subsequent time point, but the magnitude of the path coefficients decreased over time. The path coefficient from mathematics achievement in the fall of fourth grade to FNLE in the spring of fourth grade was 0.25 (a large effect; Keith, 2006). However, analogous coefficients decreased to 0.16 (a medium-to-large effect) by the end of the study (i.e., the path from spring of fifth grade achievement to winter of sixth grade FNLE.

The path coefficients from FNLE to mathematics achievement were consistently somewhat larger than those from achievement to FNLE (with the exception of the path from FNLE in fall of fifth grade to mathematics achievement later that year), but likewise decreased in strength over the course of the study. A caveat is that the path from FNLE in fall of fifth grade to spring fifth grade mathematics achievement that year is non-significant. The autoregressive paths from spring fifth grade mathematics achievement to later mathematics achievement time points are also very weak compared to other autoregressive links. One possible explanation for this is that the path from spring fifth grade

Table 2
Correlations among dependent measures.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Comp F-4th	–																
2. Comp S-4th	0.658	–															
3. Comp F-5th	0.592	0.820	–														
4. Comp S-5th	0.446	0.577	0.649	–													
5. Comp W-6th	0.417	0.545	0.599	0.861	–												
6. Comp S-6th	0.507	0.466	0.420	0.318	0.291	–											
7. FNLE W-4th	0.618	0.525	0.491	0.346	0.327	0.714	–										
8. FNLE S-4th	0.591	0.635	0.604	0.402	0.346	0.674	0.861	–									
9. FNLE F-5th	0.502	0.605	0.717	0.414	0.371	0.592	0.792	0.808	–								
10. FNLE S-5th	0.444	0.511	0.610	0.633	0.588	0.575	0.699	0.698	0.833	–							
11. FNLE W-6th	0.631	0.483	0.426	0.335	0.319	0.466	0.424	0.429	0.341	0.298	–						
12. Ach F-4th	0.488	0.444	0.416	0.347	0.320	0.657	0.621	0.656	0.611	0.618	0.459	–					
13. Ach S-4th	0.568	0.514	0.475	0.375	0.323	0.669	0.691	0.704	0.633	0.603	0.397	0.767	–				
14. Ach F-5th	0.511	0.576	0.557	0.448	0.390	0.710	0.712	0.705	0.661	0.635	0.414	0.810	0.805	–			
15. Ach S-5th	0.446	0.478	0.468	0.325	0.300	0.524	0.520	0.547	0.512	0.516	0.364	0.682	0.735	0.705	–		
16. Ach F-6th	0.447	0.423	0.390	0.514	0.387	0.728	0.629	0.630	0.564	0.581	0.301	0.786	0.818	0.801	0.754	–	
17. Ach S-6th	0.514	0.543	0.570	0.608	0.537	0.641	0.676	0.693	0.712	0.699	0.403	0.796	0.796	0.828	0.703	0.781	–

Note. All correlations significant at $p < 0.001$. Comp = fraction comparison. FNLE = fraction number line estimation. Ach = mathematics achievement. F = fall. S = spring. W = winter.

Table 3
Fit indices and likelihood ratio tests comparing nested models of the relationship between mathematics achievement and fraction number line estimation performance.

	Achievement/FNLE accuracy				
	Model 1	Model 2	Model 3a	Model 3b	Model 4
	Independent first-order autoregressive	Independent higher-order autoregressive	Higher-order autoregressive unidirectional (Ach → FNLE)	Higher-order autoregressive unidirectional (FNLE → Ach)	Higher-order autoregressive cross-lagged (FNLE ↔ Ach)
df	96	90	86	86	82
χ^2	991.854	568.097	424.707	412.098	296.939
RMSEA	0.132	0.100	0.086	0.084	0.070
AIC	55,825.356	55,413.599	55,278.209	55,265.600	55,158.441
BIC	56,223.780	55,837.728	55,719.475	55,706.866	55,616.843
CFI	0.798	0.892	0.924	0.926	0.951
TLI	0.737	0.850	0.889	0.893	0.926
Likelihood ratio test		(vs. Model 1) $\chi^2(6) = 423.76$ $p < 0.0001$	(vs. Model 2) $\chi^2(4) = 143.39$ $p < 0.0001$	(vs. Model 2) $\chi^2(4) = 156.00$ $p < 0.0001$	(vs. Model 3a) $\chi^2(4) = 115.16$ $p < 0.0001$ (vs. Model 3b) $\chi^2(4) = 127.77$ $p < 0.0001$

Note. $N = 536$. Ach = mathematics achievement. FNLE = fraction number line estimation accuracy. RMSEA = root mean square error of approximation. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. CFI = Comparative Fit Index. TLI = Tucker-Lewis Index. Standardized root mean square residual (SRMR) is not reported because of missing values.

Table 4
Background variable standardized covariances and coefficients for the FNLE/mathematics achievement model.

	Covariances							Coefficients		
	1	2	3	4	5	6	7	FNLE W-4th	Ach F-4th	Teacher effect slope S-4th
1. Age	–							− 0.001	0.088	0.184*
2. Gender	− 0.033	–						− 0.096*	− 0.039	0.262**
3. Low income	0.144**	0.008	–					− 0.063	− 0.077*	− 0.085
4. Vocabulary	− 0.322***	− 0.046	− 0.426***	–				0.278***	0.306***	− 0.206
5. Matrix reasoning	− 0.341***	0.109*	− 0.273***	0.532***	–			0.188***	0.154***	− 0.035
6. Reading	− 0.554***	− 0.012	− 0.245***	0.433***	0.321***	–		− 0.067	0.127**	0.002
7. Working memory	− 0.206***	− 0.021	− 0.088	0.280***	0.359***	0.235***	–	0.095*	0.077**	− 0.029
8. Attentive behavior	− 0.184***	0.199***	− 0.224***	0.369***	0.407***	0.407***	0.312***	0.226***	0.340***	0.052

Note. Ach = mathematics achievement. FNLE = fraction number line estimation accuracy. W-4th = winter fourth grade. F-4th = fall fourth grade. S-4th = spring fourth grade. Gender: 1 = female, 0 = male. Low income: 1 = low income, 0 = not low income.
*** $p < 0.001$.
** $p < 0.01$.
* $p < 0.05$.

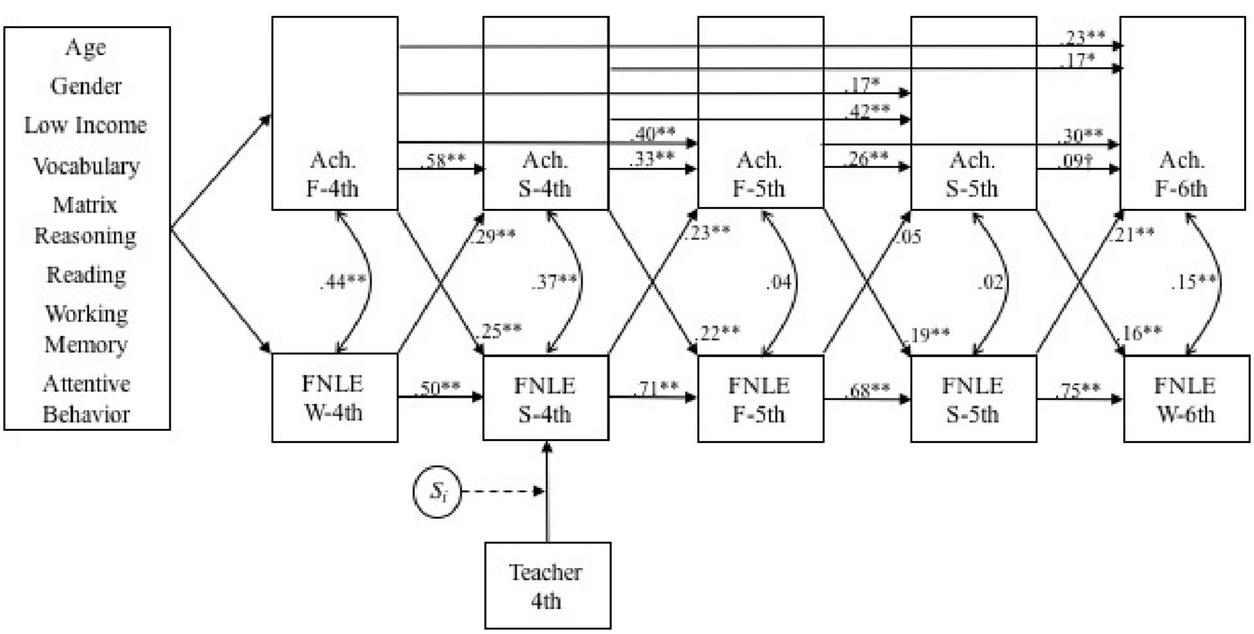


Fig. 2. Final higher-order autoregressive cross-lagged panel model (Model 4) showing relations between fraction number line estimation and mathematics achievement. Background variable covariances and independent paths not shown (see Table 4). *** $p < 0.001$. ** $p < 0.01$. * $p < 0.05$. Ach = mathematics achievement. FNLE = fraction number line estimation accuracy. F = fall. S = spring. W = winter.

math achievement to spring sixth grade achievement has a smaller weight simply because it is competing for variance against a large number of autoregressive links. Consistent with this explanation, every path between consecutive mathematics achievement measures has a smaller coefficient than all those preceding it. Alternatively, this result may suggest something distinctive about the spring fifth grade mathematics achievement measure. Because the content of state achievement test varies from year to year, it could be that the content tested that year idiosyncratically bore a weaker relationship to fractions, or perhaps the test may simply not have been as well-designed as prior or subsequent mathematics achievement measures. Nonetheless, the overall trend of directionality is clear, and specific investigation into why results for this particular time point are weaker is outside the scope of the present study.

Autoregressive paths for both the mathematics achievement and FNLE scores revealed the longitudinal stability of this measure when correlational and cross-lagged associations were simultaneously accounted for. For FNLE, autoregressive path coefficients ranged from 0.50 to 0.75, all large effects. For mathematics achievement, autoregressive path coefficients between consecutive time points ranged from a marginally significant magnitude of 0.09 to 0.58 (a large effect). Higher-order auto-regressive coefficients for mathematics achievement included several medium to large effects. That is, knowledge assessed on earlier mathematics achievement assessments showed downstream effects and strongly predicted performance on later mathematics achievement assessments.

4.2. Relations between fraction comparisons and mathematics achievement

We constructed a series of nested panel models as described previously. Similar to the previous analysis, fit indices for all intermediate models (Table 5) favor the higher-order AR-CL model (Model 4; $\chi^2(110) = 257.08$, RMSEA = 0.050, CFI = 0.967, TLI = 0.951, AIC = 65,957.144, BIC = 66,471.240). Furthermore, likelihood ratio test results indicate that including the full set of cross-lagged paths significantly improved the fit of the higher-order autoregressive model (Model 4) over that of both unidirectional models. The results indicate that relationships between fraction comparisons and mathematics achievement from fourth to sixth grade are bidirectional.

As with the model relating FNLE to achievement, we tested to see whether random slopes for effects of teacher exhibited significant variance for the AR-CL model. Just as was the case previously, the only slope with significant variances was that associated with the fractions measure (comparisons, in this case) in the spring of 4th grade, $s^2 = 18.88$, $SE = 4.73$, $p < 0.001$. Random slopes for all other scores at all other time points were non-significant (all $ps > 0.05$). Thus, the final model (Model 5) for fraction comparisons was a full AR-CL with a random slope for the effect of teacher on the comparison score in the spring of 4th grade, with this slope further regressed on the full set of covariates.

Table 6 provides covariances among background variables and path coefficients relating each of these variables to baseline scores for fraction comparison and mathematics achievement, as well as effects on the random slope for the effect of teacher in the spring of 4th grade. The only variable that uniquely predicted baseline scores for fraction comparison was income status. Paths from age, vocabulary, matrix reasoning, reading, and attentive behavior to baseline mathematics achievement were statistically significant, indicating that these variables were uniquely associated with mathematics achievement after controlling for the other associations in the model. None of the covariates significantly predicted the random slope of the effect of teacher.

Fig. 3 shows a path diagram for Model 5 with standardized coefficients for the auto-regressive cross-lagged portion of the model. We found that mathematics achievement scores from fourth to sixth grade consistently predicted fraction comparisons scores at a later time point, but the strength of this relation decreased over time. For example, the path coefficient from mathematics achievement in fall of fourth grade to fraction comparisons in spring of fourth grade was 0.43 (a large effect), while by the end of the study, the path coefficient from mathematics achievement in fall of sixth grade to fraction comparisons in winter of sixth grade was 0.04 (not significant). Therefore, although earlier mathematics achievement did relate to later fraction comparisons skill, this relationship weakened as the study progressed. We also found that fraction comparisons scores predicted mathematics achievement at most time points, with the exception of a marginally significant ($p = 0.06$) result for the path from fall of fourth grade to spring of fourth grade and a non-significant path from fall of fifth grade to spring of fifth grade. In the latter case, we note that this result is similar to that obtained for the FNLE/achievement model. Given that this is this same path was non-significant in both cases, yet the cross-lagged paths from comparison scores to achievement scores subsequently return to magnitudes comparable to prior paths, this suggests that the non-significance may be more likely to be due to unique characteristics of the achievement test in spring of fifth grade rather

Table 5
Fit indices and likelihood ratio tests comparing nested models of the relationship between mathematics achievement and fraction comparison performance.

	Achievement/Comp accuracy				
	Model 1	Model 2	Model 3a	Model 3b	Model 4
	Independent first-order autoregressive	Independent higher-order autoregressive	Higher-order autoregressive unidirectional (Ach → Comp)	Higher-order autoregressive unidirectional (Comp → Ach)	Higher-order autoregressive cross-lagged (Comp ↔ Ach)
df	130	120	115	115	110
χ^2	907.039	398.428	295.786	358.231	260.120
RMSEA	0.106	0.066	0.054	0.063	0.050
AIC	66,564.064	66,075.452	65,982.810	66,045.255	65,957.144
BIC	66,992.477	66,546.706	66,475.486	66,537.930	66,471.240
CFI	0.847	0.945	0.964	0.952	0.970
TLI	0.809	0.926	0.950	0.933	0.956
Likelihood ratio test		(vs. Model 1) $\chi^2(10) = 508.61$ $p < 0.0001$	(vs. Model 2) $\chi^2(5) = 102.64$ $p < 0.0001$	(vs. Model 2) $\chi^2(5) = 40.20$ $p < 0.0001$	(vs. Model 3a) $\chi^2(5) = 35.67$ $p < 0.0001$ (vs. Model 3b) $\chi^2(5) = 98.11$ $p < 0.0001$

Note. $N = 536$. Ach = mathematics achievement. Comp = fraction comparison. RMSEA = root mean square error of approximation. AIC = Akaike Information Criterion. BIC = Bayesian Information Criterion. CFI = Comparative Fit Index. TLI = Tucker-Lewis Index. Standardized root mean square residual (SRMR) is not reported because of missing values.

Table 6
Background variable standardized covariances and coefficients for the comparisons/mathematics achievement model.

	Covariances							Coefficients		
	1	2	3	4	5	6	7	Comp W-4th	Ach F-4th	Teacher effect slope S-4th
1. Age	–							−0.042	0.086	−0.132
2. Gender	−0.032	–						−0.080†	−0.039	0.017
3. Low income	0.143**	0.008	–					−0.123*	−0.079*	−0.044
4. Vocabulary	−0.323***	−0.047	−0.426***	–				0.153**	0.304***	0.036
5. Matrix reasoning	−0.342***	0.108*	−0.272***	0.532***	–			0.095	0.153**	−0.085
6. Reading	−0.554***	−0.013	−0.244***	0.433***	0.321***	–		0.037	0.129**	−0.075
7. Working memory	−0.207***	−0.021	−0.087	0.281***	0.360***	0.234***	–	0.051	0.079*	0.024
8. Attentive behavior	−0.184***	0.198***	−0.224***	0.369***	0.406***	0.408***	0.312***	0.085	0.339***	0.049

Note. Ach = mathematics achievement. Comp = fraction comparisons score. W-4th = winter fourth grade. F-4th = fall fourth grade. S-4th = spring fourth grade. Gender: 1 = female, 0 = male. Low income: 1 = low income, 0 = not low income.
*** $p < 0.001$.
** $p < 0.01$.
* $p < 0.05$.

than competition from an increasing number of prior autoregressive paths among achievement scores. In this model, coefficients for paths leading from comparison scores to achievement scores were relatively stable, in contrast to analogous paths for FNLE, which appeared to decrease steadily over time.

Autoregressive paths for both the fraction comparisons and mathematics achievement scores revealed the longitudinal stability of these measures when simultaneously accounting for correlational and cross-lagged associations. That is, scores on these measures at one time point were statistically significantly predictive of scores at a subsequent time point, even when controlling for the other associations in the model. For fraction comparisons, autoregressive path coefficients ranged from 0.12 (small effect) to 0.82 (large effect). For mathematics achievement, first-order autoregressive path coefficients between consecutive time points ranged from 0.17 to 0.74 (medium to large effects). Furthermore, significant higher-order autoregressive coefficients ranged from 0.16 to 0.42, representing medium-to-large effects linking mathematics achievement at one time point to mathematics achievement at subsequent, non-adjacent time points. These results suggest that mathematics achievement scores at earlier points in time have downstream effects on later achievement that are not reflected in test scores at intervening time points. This may be due to variations in focal content over time. For example, the content on the fourth-grade mathematics achievement test may not have necessarily been on the fifth-grade test; however, knowledge of this content may still have a "trickle-down" effect that contributed to scores in sixth grade, as mathematics topics are often revisited later on in schooling.

Looking across the final models for both comparisons and FNLE, the paths from fraction comparisons and FNLE, respectively, to mathematics achievement were significant at nearly all lagged time points, with the notable exception of paths from fall of fifth grade to spring of fifth grade. Further evidence the achievement test in the spring of fifth grade may be something of an outlier comes in the form of pairwise correlations between scores on this test and achievement test scores at other time points. These pairwise correlations were lower across the board (rs ranging from 0.68 to 0.74) compared to pairwise correlations among achievement test scores that did not include the spring of fifth grade test (rs ranging from 0.77 to 0.85).

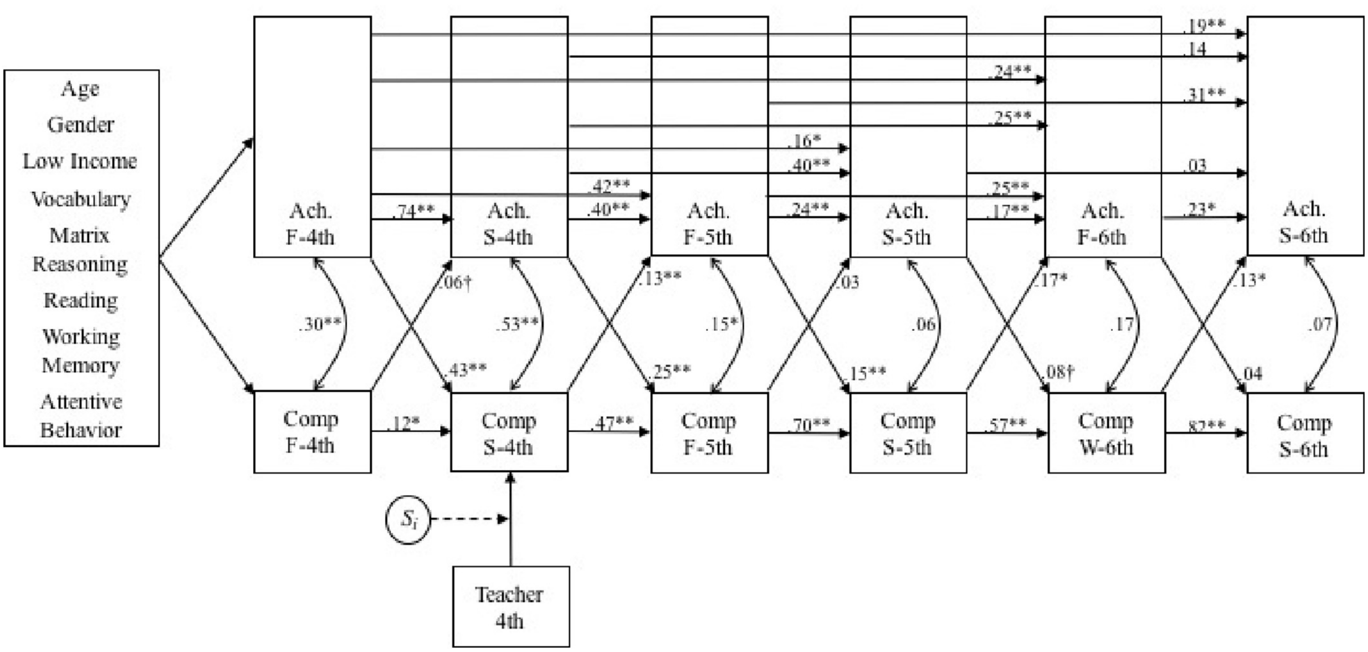


Fig. 3. Final higher-order autoregressive cross-lagged panel model (Model 4) showing relations between fraction comparison and mathematics achievement. Background variable covariances and independent paths not shown (see Table 6). *** $p < 0.001$. ** $p < 0.01$. * $p < 0.05$. † $p < 0.10$. Ach = mathematics achievement. Comp = fraction comparison accuracy. F = fall. S = spring. W = winter.

In summary, mathematics achievement scores consistently predicted both FNLE and fraction comparison scores at later time points, but the strength of the relationship decreased over time. Significant autoregressive paths for both FNLE and fraction comparison revealed the longitudinal stability of fractions and mathematics achievement measures when accounting for correlational and cross-lagged associations simultaneously. The main difference between the two final models was the strength of the relationship between each fraction measure and mathematics achievement. Fraction number line estimation accuracy and fraction comparisons were generally predictive of subsequent mathematics achievement, but the strength of the relationship (as evidenced by the magnitude of the standardized path coefficients) was considerably greater for the model relating math achievement to FNLE scores.

5. Discussion

The present study is to our knowledge the first to demonstrate a bidirectional developmental relation between fraction magnitude knowledge and mathematics achievement from fourth through sixth grade, the period during which fractions are introduced and most classroom fractions instruction occurs. We used autoregressive cross-lagged panel modeling to examine relations between overall mathematics achievement and two commonly used fraction magnitude measures (i.e., fraction comparisons and fraction number line estimation), while controlling for background and general cognitive variables. Previous research showed that fraction magnitude knowledge predicts mathematics achievement, but not the other way around (Bailey et al., 2012). Our analysis provides a more complete developmental picture by using multiple measures at a greater number of time points. Our finding of a bidirectional relation between mathematics achievement and each fraction measure supports our hypothesis that the fraction magnitude understanding and broad mathematics achievement influence one another reciprocally. That is, early mathematics achievement likely provides a foundation upon which to learn fractions, and strong fraction magnitude understanding, in turn, bolsters general mathematics achievement later in school.

The bidirectional relation between FNLE and broad mathematics achievement was consistently strong over the three grades assessed, weakening only slightly over time. Our study extends the work of Friso-van den Bos et al. (2015), who found that whole number line estimation skill and mathematics achievement mutually supported each other early in development, highlighting the continuous role of numerical magnitude knowledge in learning increasingly complex mathematics. Taken together, the findings support the integrated theory of numerical development (Siegler & Lortie-Forgues, 2014), which posits that number learning is a progressive broadening of the set of numbers whose magnitudes can be accurately represented. When acquiring more advanced mathematics knowledge (e.g., decimals and percentages), students further refine their understanding of magnitude, which then builds general mathematics skills (e.g., pre-algebra; Siegler et al., 2011; Siegler & Lortie-Forgues, 2014). Even though most fraction instruction occurs in fourth and fifth grade, according to common mathematics curriculum benchmarks, students may continue to generate new strategies for locating fractions on the number line as their mathematics learning continues (Siegler & Lortie-Forgues, 2014). Indeed, sixth- and eighth-grade students report transforming fractions into different forms, such as percentages, when locating them on the number line (Siegler et al., 2011; Siegler & Lortie-Forgues, 2014). In other words, transfer may occur in both directions (e.g., Camarata, Nelson, Gillum, & Camarata, 2009; Hohensee, 2011), such that learning more advanced mathematics allows students to deepen their understanding of fractions concepts taught earlier and learning about fractions improves later achievement in advanced mathematics.

The FNLE task and the fraction comparison task both yielded bidirectional relations with mathematics achievement that weakened over time, suggesting that both tasks are functioning in a similar way. However, the results of our CFA suggest that FNLE may be a purer and more sensitive measure of magnitude understanding. The two tasks have notable differences in what they assess as well as the strategies that they elicit. The fraction number line estimation task encourages students to map numerical magnitudes on one continuous dimension, and also may tap into proportional reasoning (Schneider, Thompson, & Rittle-Johnson, in press). Because the FNLE task uses a continuous scale, students are unlikely to ever reach ceiling, or perfect accuracy, on the task. As mentioned previously, their FNLE acuity may improve even after formal fraction instruction ends, as students continue to expand their knowledge of numerical magnitudes. It is possible that some students used strategies other than mentally representing the magnitude of a given fraction to solve the task (e.g., they could mentally divide the number line into the number of pieces as indicated by the denominator, then count as many pieces as indicated by the numerator). As noted above, Siegler et al. (2011) found that older students in sixth and eighth grades used line segmentation or numerical transformation (rounding, simplifying, translating to a percent or decimal) at least some of the time. However, we do not believe that segmentation or transformation strategies were applied consistently across all items, since several of the fractions estimated (e.g., 2/13, 1/19, 7/9) do not lend themselves to efficient and accurate use of these strategies by young students.

The fraction comparison task, on the other hand, is binary, only assesses whether students can make larger/smaller judgments (Schneider et al., in press), and may be more likely to encourage strategies that are largely procedural (though this is likely dependent on the particular fractions that are presented as well as the order they are presented in; Schneider et al., in press). As such, FNLE and fraction comparison measures are not fully interchangeable assessments of fraction magnitude understanding, despite the fact that they have previously been treated as such in the research literature. Rather, researchers should carefully consider both types assessments to determine whether they are aligned with the research question of interest. For example, the FNLE task might be more appropriate when trying to assess magnitude understanding with greater sensitivity to growth or long-term effects on achievement, while the comparison task may be better used with the understanding that multiple strategies (including, but not limited to magnitude understanding) can be used to solve the items. Further, fraction comparison tasks may more closely reflect the kinds of problems found in school curricula and everyday life.

Bailey et al. (2012) found that mathematics achievement in sixth grade does not predict fraction magnitude knowledge (using a fraction comparisons task) in seventh grade. Although we found that mathematics achievement predicted fraction magnitude knowledge over the course of our study, it should be noted that the relation diminished over time, especially for comparisons (a task that is similar to the fraction measure used by Bailey and colleagues). As discussed previously, this weakened association may be a function of the nature of the fractions comparisons measure. Comparisons assess students' understanding of a few key aspects of fractions (for example, the inverse relation between the magnitude of the fraction and the size of the denominator) and some items can be completed using simple procedural strategies.

Although it was not the primary aim of the present study, we were able to examine the lasting influence of earlier test performance on performance at subsequent time points by modeling autoregressive effects. We found significant medium-to-large effects for all measures, indicating the presence of considerable longitudinal stability for fraction comparisons and FNLE, as well as mathematics achievement. Scores on these measures predicted scores on the same measure at subsequent time points, even after controlling for a variety of other background measures in our models. It is important to point out that we did not include higher-order autoregressive paths for the FNLE and fraction comparison measures, since the task items were identical every time. However, we did include higher-order autoregressive paths for the mathematics achievement measure, since the content on the test changed yearly to match grade-level appropriate benchmarks. The findings indicate that individual mathematical concepts are not learned in isolation, but rather are built upon increasingly complex knowledge structures. This stability highlights the complex trajectory of mathematics learning. Although there are mutually reinforcing relationships between fraction magnitude knowledge and general mathematics achievement, students' past performance in either area heavily influences their future performance (Hecht & Vagi, 2010).

Finally, we also found interesting effects of teacher as well as other covariates. First, it is notable that random slopes for the effect of teacher exhibited significant variance only for the fractions measure at the end of fourth grade. Given that fourth grade is when fractions are typically first introduced, this suggests that quality of instruction may be uniquely important for initial fractions learning. Second, a number of covariate effects were found. While most covariates were related to general mathematics achievement, only a subset were related to fractions understanding, most notably vocabulary, matrix reasoning, and attentive behavior. Meanwhile, demographic factors appeared to amplify (or mitigate) effects of teacher on fractions learning, at least for FNLE. Students tend to be older than their peers in cases of late school entry or retention, signals of early academic struggles. Thus, the fact that older age increases the slope of teacher effects suggests that quality of instruction is particularly important for these students. This also appears to be true for females; previous research has shown that females are more prone to math anxiety transmitted from their math anxious teachers (Beilock, Gunderson, Ramirez, & Levine, 2010). Although the teacher and covariate effects we observed were not of primary interest for our study, our findings indicate that such effects are worthy of further research.

5.1. Limitations and future directions

Although the present study primarily investigated relations between fraction magnitude knowledge and mathematics achievement, while controlling for a range of variables associated with learning, no analysis can be entirely comprehensive. For example, we did not directly assess specific instructional factors that might impact learning in this study. All of our participants were taught in schools that used common benchmarks derived from the CCSS in mathematics; however, fraction instruction does differ across fourth through sixth grades (for example, the fourth grade standards focus on topics such as comparison, equivalence, and additional and subtraction of fractions with like denominators, while the sixth grade standards focus on dividing fraction by fractions). Despite this, instructional variation is likely, particularly because each of these mathematics topics could be taught in a more conceptual way or a more procedural way. For example, fraction comparisons could be taught procedurally (cross-multiplying) or more conceptually (using benchmarks of 0, 1/2, and 1 on a number line). As such, the present study provides important evidence about the nature of fraction development within everyday instructional settings; however, future work should examine specific school- or classroom-level instructional factors that might impact the relation between fraction magnitude knowledge and mathematics achievement.

Cross-lagged panel modeling provides better evidence of directionality than other correlational and mediation designs. Our study, which provides a unique developmental account using large-scale longitudinal data during a key instructional time period, is in line with other recent work showing bidirectional relationships between fraction arithmetic skill and fraction magnitude understanding over fourth through sixth grade (Bailey, Hansen, & Jordan, 2017). However, no studies based on correlational data demonstrate clear causal connections, making future experimental work in this area necessary. It would be useful to investigate whether some specific aspects of fraction instruction influence other mathematics skills (and vice versa) more strongly than others. For example, in fourth grade, students learn about factors and multiples, which could in turn bolster their understanding of fractions (e.g., understanding the multiplicative relationship between halves and wholes, finding common denominators when solving fraction arithmetic, etc.).

We also acknowledge limitations surrounding the fraction comparison task. In the present study, we posit that the fraction comparison task might be tapping in part into rote procedural skills in addition to true magnitude understanding. First, although the use of cross-multiplication is always a possibility for older students, the time limited nature of our task makes the use of such strategies for all items implausible. We also aimed to present comparisons that could be solved in a variety of ways. For example, fractions with the same denominator can be compared by looking at the numerator and picking the larger of the two. However, other items may require more evaluation and comparison of fraction magnitudes, since they do not allow for a simple comparison of component parts. Although our comparisons task was based directly on previous research (Bailey et al., 2012) to draw conclusions across studies, future work in this area should target more items that tap into magnitude understanding and are not easily solved by quick rote strategies or simply comparing component parts of fractions (e.g., 13/50 vs. 9/60).

Finally, we acknowledge potential overlap between content on the fractions and mathematics achievement measures. Although the mathematics achievement measures assessed a wide range of topics (e.g., algebraic reasoning, whole number skills, geometry, data analysis), fractions are an important part of mathematics achievement in fourth through sixth grade. Fraction number line estimation, however, is less likely than fractions comparisons to be directly incorporated into instruction. Future work might avoid this issue by using a grade-appropriate mathematics achievement measure that does not contain questions related to fractions. In the present study, this type of analysis was not possible. Our mathematics outcome measure was a state-level test administered by the school district, with scores that did not allow for examination by individual subskills.

In conclusion, our study reveals a bidirectional developmental relationship between fraction magnitude knowledge and mathematics achievement between fourth through sixth grade, supporting the hypothesis that fraction magnitude understanding and broader mathematics skills mutually support one another. This finding extends previously observed relationships between whole number line estimation and mathematics achievement in earlier grades (Friso-van den Bos et al., 2015), suggesting that understanding of numerical magnitudes in general and broad mathematics achievement share a reciprocal predictive relationship. Such findings are also consistent with the central role of magnitude understanding in the integrated theory of numerical development (Siegler et al., 2011; Siegler & Lortie-Forgues, 2014). Finally, our study highlights the importance of continuing to identify foundational mathematics skills that may iteratively support the development of fraction magnitude understanding (e.g., Bailey et al., 2014; Resnick et al., 2016) and of teaching these core skills while concurrently working to build fraction magnitude understanding.

References

American Institutes for Research (2012). DCAS 2011-2012 technical report. Retrieved from http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/111/Vol1_Annual_TechRep.pdf.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bailey, D. H., Hansen, N., & Jordan, N. C. (2017). The codevelopment of children's fraction arithmetic skill and fraction magnitude understanding. *Journal of Educational Psychology, 109*(4), 509–519. http://dx.doi.org/10.1037/edu0000152.

Bailey, D. H., Hoard, M. K., Nugent, L., & Geary, D. C. (2012). Competence with fractions predicts gains in mathematics achievement. *Journal of Experimental Child Psychology, 113*, 447–455. http://dx.doi.org/10.1016/j.jecp.2012.06.004.

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science, 17*, 775–785. http://dx.doi.org/10.1111/desc.12155.

Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychological*

Science, 25, 2017–2026. http://dx.doi.org/10.1177/0956797614547539.

Bailey, D. H., Zhou, X., Zhang, Y., Cui, J., Fuchs, L. S., Jordan, N. C., ... Siegler, R. S. (2015). Development of fraction concepts and procedures in U.S. and Chinese children. *Journal of Experimental Child Psychology, 129*, 68–83. http://dx.doi.org/10.1016/j.jecp.2014.08.006.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107*(5), 1860–1863. http://dx.doi.org/10.1073/pnas.0910967107.

Bentler, P. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107*, 238–246. http://dx.doi.org/10.1037/0033-2909.107.2.238.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606. http://dx.doi.org/10.1037/0033-2909.88.3.588.

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: rethinking a developmental workhorse. *Child Development, 88*, 1186–1206. http://dx.doi.org/10.1111/cdev.12660.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach.* Hoboken, NJ: Wiley.

Booth, J. L., & Newton, K. J. (2012). Fractions: could they really be the gatekeeper's doorman? *Contemporary Educational Psychology, 37*(4), 247–253. http://dx.doi.org/10.1016/j.cedpsych.2012.07.001.

Booth, J. L., Newton, K. J., & Twiss-Garrity, L. K. (2014). The impact of fraction magnitude knowledge on algebra performance and learning. *Journal of Experimental Child Psychology, 118*, 110–118. http://dx.doi.org/10.1016/j.jecp.2013.09.001.

Brown, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. Long (Eds.). *Testing structural equation models* (pp. 445–455). Newbury Park, CA: Sage.

Burkholder, G. J., & Harlow, L. L. (2003). An illustration of a longitudinal cross-lagged design for larger structural equation models. *Structural Equation Modeling, 10*(3), 465–486. http://dx.doi.org/10.1207/S15328007SEM1003_8.

Camarata, S., Nelson, K. E., Gillum, H., & Camarata, M. (2009). Incidental receptive language growth associated with expressive grammar intervention in SLI. *First Language, 29*, 51–63. http://dx.doi.org/10.1177/0142723708098810.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

de Jonge, J., Dormann, C., Janssen, P. P. M., Dollard, M. F., Landeweerd, J. A., & Nijhuis, F. J. N. (2001). Testing reciprocal relationships between job characteristics and psychological well-being: A cross-lagged structural equation model. *Journal of Occupational and Organizational Psychology, 74*, 29–46. http://dx.doi.org/10.1348/096317901167217.

Delaware Department of Education (2013). DCAS subjects and benchmarks. Retrieved from http://de.portal.airast.org/wpcontent/uploads/2013/06/DCAS_Subjects_and_Benchmarks.pdf.

DeWolf, M., & Vosniadou, S. (2015). The representation of the fraction magnitudes and the whole number bias reconsidered. *Learning and Instruction, 37*, 39–49. http://dx.doi.org/10.1016/j.learninstruc.2014.07.002.

Dunn, L., & Dunn, D. (2007). *PPVT-4, peabody picture vocabulary test manual* (4th ed.). Minneapolis, MN: Pearson Assessment.

Farrell, A. D. (1994). Structural equation modeling with longitudinal data: Strategies for examining group differences and reciprocal relationships. *Journal of Consulting and Clinical Psychology, 62*(3), 477–487. http://dx.doi.org/10.1037/0022-006X.62.3.477.

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology, 123*(1), 53–72. http://dx.doi.org/10.1016/j.jecp.2014.01.013.

Fazio, L. K., DeWolf, M., & Siegler, R. S. (2016). Strategy use and strategy choice in fraction magnitude comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1–16. http://dx.doi.org/10.1037/xlm0000153.

Finkel, S. (1995). *Causal analysis with panel data.* London: Sage.

Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., & Van Lieshout, E. C. D. M. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology, 134*, 12–29. http://dx.doi.org/10.1016/j.jecp.2015.02.002.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., ... Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*(1), 29–43. http://dx.doi.org/10.1037/0022-0663.98.1.29.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., & Bryant, J. D. (2010). Do different types of school mathematics depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology, 46*(6), 1731–1746. http://dx.doi.org/10.1037/a0020662.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*, 4–15. http://dx.doi.org/10.1177/00222194040370010201.

Geary, D. C. (2011). Cognitive predictors of individual differences in achievement growth in mathematics: A five-year longitudinal study. *Developmental Psychology, 47*, 1539–1552. http://dx.doi.org/10.1037/a0025510.

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology, 48*(5), 1229–1241. http://dx.doi.org/10.1037/a0027433.

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in nonverbal number acuity correlate with maths achievement. *Nature, 455*, 665–668. http://dx.doi.org/10.1038/nature07246.

Hansen, N., Jordan, N. C., Siegler, R. S., Fernandez, E., Gersten, R., Fuchs, L., & Micklos, D. (2015). General and math-specific predictors of sixth-graders' knowledge of fractions. *Cognitive Development, 35*, 34–49. http://dx.doi.org/10.1016/j.cogdev.2015.02.011.

Hecht, S., Close, L., & Santisi, M. (2003). Sources of individual differences in fraction skills. *Journal of Experimental Child Psychology, 86*(4), 277–302. http://dx.doi.org/10.1016/j.jecp.2003.08.003.

Hecht, S. A., & Vagi, K. J. (2010). Sources of group and individual differences in emerging fraction skills. *Journal of Educational Psychology, 102*, 843–859. http://dx.doi.org/10.1037/a0019824.

Hohensee, C. (2011). *Backward transfer: How mathematical understanding changes as one builds upon it.* (Unpublished dissertation).

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology, 116*(1), 45–58. http://dx.doi.org/10.1016/j.jecp.2013.02.001.

Jordan, N. C., Resnick, I., Rodrigues, J., Hansen, N., & Dyson, N. (2016). The Delaware longitudinal study of fraction learning: Implications for helping children with mathematics learning difficulties. *Journal of Learning Disabilities*. http://dx.doi.org/10.1177/0022219416662033 (Special Issue: Fraction Learning). Advance online publication.

Keith, T. Z. (2006). *Multiple regression and beyond.* Boston: Allyn & Bacon.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: The Guilford Press.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90*, 1112–1121. http://dx.doi.org/10.1080/01621459.1995.10476615.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4*, 192–211. http://dx.doi.org/10.1037/1082-989X.4.2.192.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). Common core state standards for mathematics. Common Core State Standards Initiative. Retrieved from http://www.corestandards.org/assets/CCSSI_Mathematics%20Standards.pdf.

Pickering, S., & Gathercole, S. (2001). *Working memory test battery for children.* London: The Psychological Corporation.

Resnick, I., Jordan, N. C., Hansen, N., Rajan, V., Rodrigues, J., Siegler, R. S., & Fuchs, L. S. (2016). Developmental growth trajectories in understanding of fraction magnitude from fourth through sixth grade. *Developmental Psychology, 52*(5), 746–757. http://dx.doi.org/10.1037/dev0000102.

Rinne, L. F., Ye, A., & Jordan, N. C. (2017). Development of fraction comparison strategies: A latent transition analysis. *Developmental Psychology, 53*(4), 713–730. http://dx.doi.org/10.1037/dev0000275.

Schneider, M., Thompson, C. A., & Rittle-Johnson, B. (2017). Associations of magnitude comparison and number line estimation with mathematical competence: A comparative review. To appear in In P. Lemaire (Ed.). *Cognitive development from a strategy perspective: A festschrift for Robert S. Siegler*London: Psychology Press (in press).

Seethaler, P. M., Fuchs, L. S., Star, J. R., & Bryant, J. R. (2011). The cognitive predictors of computational skill with whole versus rational numbers: An exploratory study. *Learning and Individual Differences, 21*, 536–542. http://dx.doi.org/10.1016/j.lindif.2011.05.002.

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.). *Handbook of developmental research methods* (pp. 265–278). New York, NY: Guilford Press.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428–444. http://dx.doi.org/10.1111/j.1467-8624.2004.00684.x.

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., ... Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science, 23*(7), 691–697. http://dx.doi.org/10.1177/0956797612440101.

Siegler, R. S., & Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Development Perspectives, 8*(3), 144–150. http://dx.doi.org/10.1111/cdep.12077.

Siegler, R. S., & Pyke, A. A. (2013). Developmental and individual differences in understanding of fractions. *Developmental Psychology, 49*(10), 1994–2004. http://dx.doi.org/10.1037/a0031200.

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*(4), 273–296. http://dx.doi.org/10.1016/j.cogpsych.2011.03.001.

StataCorp (2013). *Stata statistical software: Release 13.* College Station, TX: StataCorp LP.

Swanson, H. L. (2011). Working memory, attention, and mathematical problem solving: A longitudinal study of elementary school children. *Journal of Educational Psychology, 103*(4), 821–837. http://dx.doi.org/10.1037/a0025114.

Swanson, J., Schuck, S., Mann, M., Carlson, C., Hartman, K., & Sergeant, J. (2006). *Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and SWAN rating scales. Unpublished manuscript.* Irvine: University of California.

Torbeyns, J., Schneider, M., Xin, Z., & Siegler, R. S. (2015). Bridging the gap: Fraction understanding is central to mathematics achievement in students from three different continents. *Learning and Instruction, 37*, 5–13. http://dx.doi.org/10.1016/j.learninstruc. 2014.03.002.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *TOWRE: Test of word reading*

efficiency. Austin, TX: Pro-Ed.
Tucker, L., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10. http://dx.doi.org/10.1007/BF02291170.
Vukovic, R. K., Fuchs, L. S., Geary, D. S., Jordan, N. C., Gersten, R., & Siegler, R. S. (2014). Sources of individual differences in children's understanding of fractions. *Child Development, 85*(4), 1461–1476. http://dx.doi.org/10.1111/cdev.12218.
Watts, T. W., Duncan, G. J., Chen, M., Claessens, A., Davis-Kean, P. E., Duckworth, K., ... Susperreguy, M. I. (2015). The role of mediators in the development of longitudinal mathematics achievement associations. *Child Development, 86*(6), 1892–1907. http://dx.doi.org/10.1111/cdev.12416.
Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence (WASI)*. San Antonio, TX: Harcourt Assessment.
Zapf, D., Dormann, C., & Frese, M. (1996). Longitudinal studies in organizational stress research: A review of the literature with reference to methodological issues. *Journal of Occupational Health Psychology, 1*(2), 145–169. http://dx.doi.org/10.1037/1076-8998.1.2.145.