Tracking fraction knowledge development using grade-appropriate assessments

Isabella Starling-Alves[1], Xiangyi Liao[2], Qi Huang[2],

Daniel M. Bolt[2], Edward M. Hubbard[2], & Percival G. Matthews[2]

1.Department of Psychology and Human Development, Vanderbilt University

2. Department of Educational Psychology, University of Wisconsin Madison

Corresponding author: Isabella Starling Alves; isabella.starling.alves@vanderbilt.edu

The Number Lab

Department of Psychology and Human Development

Hobbs Hall, 1818 DeFord Bailey

Vanderbilt University

Nashville, TN 37212

Acknowledgements: This research was supported by a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD088585) to E.M.H. and P.G.M. and by a core grant (U54-HD090256-01) to the Waisman Intellectual and Developmental Disabilities Research Center. I.S.A. was supported by CAPES (Doc-Pleno, 88881.128282/2016-01). The authors thank Mark R. Lewis, Elizabeth Y. Toomarian, John V. Binzak, Priya B. Kalra, Yunji Park, and Alexandria A. Viegut for their input on the design of the FKA. We also thank Jennifer C. Hathaway, Grace C. George, Mia Prifti, and Ashley L. Ezpeleta and all of the undergraduate research assistants on the LAMBDA team for their critical contributions to collecting this data. Finally, we thank the families for their long-term participation in this study.

Author contributions: E.M.H. and P.G.M. designed the study and supervised every stage of the research. I.S.A. was part of the team that collected the data and wrote the first version of the manuscript. I.S.A., X.L., and Q.H. conducted the analyses, and D.M.B. supervised it. All authors interpreted the results and critically evaluated and edited the manuscript. All authors approved the final draft for publication.

Competing interests: Authors have no competing interests to disclose.

Abstract

Fraction knowledge predicts higher-order mathematics. Valid measures are crucial for understanding its developmental trajectory. This accelerated longitudinal study designed and validated grade-appropriate versions of the Fraction Knowledge Assessment (FKA) and examined children's development from 2nd to 6th grade. Methods: We measured the fraction knowledge of 342 children. We assessed FKAs reliability, validity, and factor structure and used Item Response Theory (IRT) to calibrate item difficulty and link across versions. Then, we investigated longitudinal changes in children’s fraction knowledge. Results and Conclusion: All FKAs had good internal consistency and concurrent validity, and fit well with a one-dimensional model. IRT analyses revealed a shift from fraction magnitude items discriminating students in early grades, to operation items discriminating students in advanced grades. Development was more pronounced from 2nd to 3rd than from 5th to 6th grade, aligning with local curricula. We share the FKA versions to support further research and educational practice

Keywords: fraction knowledge, fraction knowledge assessment, rational number development

Fraction knowledge is one of the main predictors of advanced mathematics skills and may influence academic, occupational, and health outcomes (Bailey et al., 2012; National Mathematics Advisory Panel, 2008). However, fractions are one of the most challenging topics in K-12 mathematics (NAEP, 2017). Children and adults encounter difficulties grasping the magnitude of fractions, executing fraction operations, and applying fraction concepts to address academic and real-world problems (Neagoy, 2017; Stigler et al., 2010). Given the importance and difficulty of fractions, it is imperative to monitor the acquisition of fraction knowledge and identify how students respond to fractions training, which can only be achieved with valid and reliable measures of fraction knowledge.

We argue that three intertwined issues have hampered the validity of fraction knowledge measurement and, consequently, our understanding of its development: 1) limited information regarding the psychometric characteristics of fraction knowledge measures; 2) insufficient attention given to calibrating and linking various measures across developmental stages; and 3) the lack of freely shared, standardized measures investigating formal fraction knowledge in the elementary school years. To address this gap, we designed and tested a novel assessment—the Fraction Knowledge Assessment (FKA)—which is applicable for use in 2nd, 3rd, 5th, and 6th grades. We provided evidence for FKA validity using both classical test theory (CTT) and item response theory (IRT). Subsequently, we linked different versions of the FKA through IRT scaling and Maximum Likelihood Estimation (MLE). Finally, we investigated developmental shifts in students' fraction knowledge from 2nd to 3rd and 5th to 6th grade.

The "fraction knowledge" construct has been broadly defined in academic literature. Many researchers have defined it as a blend of conceptual and procedural knowledge (Bailey et al., 2014; Booth & Newton, 2012; Hansen et al., 2015; Lenz et al., 2020). Following Bailey and colleagues (2015), we define *conceptua*l fraction knowledge as the capacity to comprehend fraction facts and properties, encompassing elements such as fraction notation, magnitude, equivalence, and density. We define procedural fraction knowledge as proficiency in applying a series of algorithms to solve fraction problems, particularly arithmetic operations. Fraction knowledge thus encompasses a spectrum of competencies necessary for comprehending fraction magnitudes and properties, as well as for solving diverse problems involving fractions (Bailkey et al., 2015; Rittle-Johnson et al., 2001; Rittle-Johnson & Schneider, 2015).

The broad scope of the fraction knowledge construct presents challenges in both conceptualization and operationalization within developmental investigations. Some developmental studies have confined the fraction knowledge construct to a specific set of skills, such as fraction operations or magnitude comparisons (e.g., Bailey et al., 2014; Fazio et al., 2016). Others have adopted a more comprehensive approach by combining various skills—such as fraction magnitude processing, property identification, and operations—into a supposedly one-dimensional task (e.g., Matthews et al., 2016; Schneider & Stern, 2010), or a task comprising both conceptual and procedural factors (e.g., Jordan et al., 2013; Ye et al., 2016). This lack of consensus regarding the precise definition of fraction knowledge and the content to be encompassed within fraction knowledge measures makes it difficult to assess the validity of existing measures.

Understanding how conceptual and procedural knowledge are separated or integrated becomes critical to measuring fraction knowledge. Previous studies exploring this distinction have yielded mixed findings. For example, some confirmatory factor analysis studies found evidence that one- and two-dimensional factor solutions fit fraction knowledge tasks equally well (Schneider & Stern, 2010). On the contrary, other studies have identified a separation between conceptual and procedural factors within fraction knowledge tasks (e.g., Lenz et al., 2020). Beyond factor analyses, the differentiation between conceptual and procedural fraction knowledge has been supported due to their distinct relationships with cognitive skills (Ye et al., 2016). However, it is noteworthy that conceptual and procedural fraction knowledge were strongly correlated, with coefficients exceeding $r > .80$ (Hecht et al., 2003; Hallet et al., 2010; Ye et al., 2016), suggesting a potential shared underlying mechanism or even the possibility that they constitute the same construct. Consequently, conducting factor analyses is imperative for disentangling conceptual and procedural fraction knowledge and understanding the psychometric properties of current measures.

Limited psychometric information on fraction knowledge measures

Most studies of fraction knowledge development have provided limited details regarding the psychometric properties of their measurement tools. Some studies have only reported information concerning the internal consistency of their measures, primarily evaluated using Cronbach's alpha (e.g., Fazio et al., 2014; Hansen et al., 2015; Hecht & Vagi, 2012) or inter-rater reliability (e.g., Lee, 2011). Others have explored the predictive validity of their measures—for instance, Rodrigues et al. (2019), employed a receiver operating characteristic (ROC) curve analysis to demonstrate that their conceptual fraction knowledge measure identified 3rd graders at risk for developing mathematics difficulties in 4th to 6th grade. Although some studies have used IRT to evaluate the psychometric properties of their fraction knowledge measures, these investigations have mainly targeted students in advanced grades or adults (e.g., Chan et al., 2007; Dorri & Rafiepour, 2018; Ölmez & Izsák, 2021). Van Hoof and colleagues (2015) used IRT to examine a fraction knowledge measure among 4th grade students, but their measure mainly focused on students' ability to inhibit whole number information when solving fraction problems and did not include other conceptual or procedural items.

The limited psychometric examinations conducted on fraction measures complicates the interpretation and generalization of prior studies, making it challenging to deepen our comprehension of fraction knowledge development. An iterative method of task construction that involves designing and grading items based on the extant literature, employing IRT analyses to assess the items' properties, and subsequent refinement and adaptation of the items offers more in-depth insights into students' knowledge progression than traditional approaches that rank students based only on raw task scores (Rittle-Johnson et al., 2011).

Insufficient attention to calibrating and linking measures across development

During the elementary school years, students' fraction knowledge dramatically increases as they receive formal instruction on this subject. In the United States, the Mathematics Common Core Standards recommends that formal fractions instruction begin in the 3rd grade (National Governors Association, 2010). This instruction primarily emphasizes fundamental skills such as comparing fraction magnitudes and transcoding across notations (i.e., reading and writing fractions). Fractions remain a substantial component of the mathematics curriculum through the 7th and 8th grades, when students receive instruction on how to solve complex arithmetic and pre-algebra problems with fractions. It is critical to account for this progression in the mathematics curriculum to accurately measure the development of fraction knowledge in a reliable and valid manner. However, most studies focusing on fraction knowledge development have predominantly utilized used investigator-created assessments (e.g., McMullen et al., 2015; Siegler et al., 2011) or parts of omnibus standardized tests (e.g., Jordan et al., 2017; Rodrigues et al., 2019). Typically, these measures lack grade-appropriate subtests with comparable psychometric properties and do not support flexible item banking for use across various grade levels. Consequently, existing measures suffer two limitations: either a measure is employed uniformly across grades without adaptations, or grade-custom-made measures are used without undergoing rigorous psychometric testing and linkage across grades.

The sensitivity of a task will vary depending on the student's grade level. Accordingly, the use of a uniform measure across grades is susceptible to floor and ceiling effects. For example, solving mixed number division problems might be appropriate for 6^{th}-grade students but lead to floor effects for 3^{rd} graders. Conversely, transcoding fractions from verbal to written notation might be appropriate for 3^{rd} graders but lead to ceiling effects in 6^{th} graders. Finally, because the mathematics curriculum may change over time (National Governors Association, 2010), it is crucial to update and refine our measures to ensure they continue to allow unbiased grade-appropriate assessments. An alternative approach to investigating students' fraction knowledge development involves using different measures for each grade. However, when using this approach, it is essential to ascertain that increases in raw scores with grade genuinely reflect enhanced fraction knowledge, rather than differences in measure difficulty. A critical aspect of this approach is the calibration between participants' skills and the difficulty levels of the tasks to ensure fair comparison of performance across different grades, thereby minimizing bias.

Numerical cognition studies have traditionally indexed participants' fraction knowledge based on raw accuracy (e.g., Bailey et al., 2014; Hansen et al., 2015) or the percentage of correct responses in a task (e.g., Hecht et al., 2003; Kalra et al., 2020). However, raw scores depend not only on participants' ability but also on the difficulty of the items (Baker, 2001). Measures that are flexibly adapted to different grade-level proficiencies can be used in developmental studies when they incorporate cloned items that have been tested and linked via IRT (Lee & Ban, 2009). This approach enables the determination of the probability that a student with a particular ability level (i.e., ability theta, θ) will successfully complete an item of a certain difficulty level (i.e., difficulty level delta, δ), thus providing a calibration match between the abilities of the participants and the difficulty levels of the items.

Measuring fraction knowledge in early elementary school

Children may begin to develop a foundational understanding of fraction knowledge even before formal introduction to fractions in the classroom. Preschool students demonstrate the ability to compare and solve operations involving nonsymbolic ratio magnitudes, identify and estimate the magnitude of "one-half," and recognize some frequently used fractions (Gelman et al., 1991; Hurst & Cordes, 2018; Mix et al., 1999; Park et al., 2021). Young children's fraction knowledge repertoire develops both through informal experiences at home with their caregivers (Eason & Ramani, 2020) and preschool experiences encompassing various mathematical concepts, including whole-number arithmetic, financial literacy, and geometry (Bailey et al., 2014; Viegut et al., 2023). These studies targeting preschool students and those below the 3rd grade indicate that the developmental trajectory of fraction knowledge starts early, a phenomenon that warrants further investigation.

Despite evidence that fraction knowledge may start developing as early as preschool, most studies on fraction knowledge development have prioritized grades where the mathematics curriculum explicitly addresses fractions, typically ranging from 3rd to 7th grade. For example, the Delaware Longitudinal Study of Fraction Learning, one of the largest endeavors in this field, followed children from 3rd to 6th grade (Jordan et al., 2017). Another developmental study, which examined fraction and decimal magnitude comparison and number line estimation, tracked children from 4th to 12th grade (Wang & Siegler, 2023). Thus, our understanding of fraction knowledge among students in early grades of elementary school and our ability to assess it effectively have been limited by the scarcity of studies targeting earlier grades.

The Present Study

In the current study, we aimed to address the absence of valid and reliable measures capable of longitudinally evaluating students' fraction knowledge across grades, commencing from early elementary school. We developed and tested the psychometric properties of a novel Fraction Knowledge Assessment (FKA) with versions tailored for 2nd, 3rd, 5th, and 6th grades. Each FKA version included items designed to assess a wide spectrum of fraction skills, encompassing both conceptual and procedural knowledge. These items were tied to the Mathematics Common Core Standards and were sourced from previous research articles, as well as from international, national, and state achievement exams (e.g., Carpenter, 1981; Hallett et al., 2012), including editions of the National Assessment of Educational Progress (NAEP) and the Trends in International Mathematics and Science Study (TIMSS).

We examined the psychometric properties of the FKA using several approaches. We used CTT to evaluate the reliability and concurrent validity of the FKA. Additionally, we used Confirmatory Factor Analysis (CFA) to examine the factor structure of the assessment. Subsequently, we employed IRT to (1) Measure students’ fraction ability while simultaneously calibrating for item difficulty on an interval scale, thereby facilitating more precise assessment, (2) Equate participants’ scores within and between grades, allowing for flexibly measuring fraction knowledge in longitudinal analyses, and (3) Compare students’ fraction knowledge across grades on a common metric, thereby enhancing the interpretability and comparability of results across different grade levels. After collecting validity evidence for the FKA, we examined participants' developmental shifts in fraction knowledge from 2nd grade, to 3rd grade, and from 5th to 6th grades, following several years of classroom fraction instruction. We then characterized students' primary challenges in fraction knowledge in each grade and examined longitudinal improvements in their fraction achievement.

Methods

Participants

The sample for the current study was drawn from a larger longitudinal investigation, which comprised participants recruited in two enrollment waves through invitation letters sent to schools in a medium-sized Midwestern city. We followed two cohorts of students in a accelerated longitudinal design: the *Younger* cohort was followed from 2nd to 5th grade, and the Older cohort was followed from 5th to 8th grade. A total of 342 participants completed grade-appropriate versions of the FKA. Younger participants took the FKAs in 2nd, 3rd, and 5th grades, whereas *Older* participants completed FKAs in 5th and 6th grades. We initially planned for these participants to also complete an FKA version in 8th grade; however, data collection was disrupted due to the COVID-19 pandemic. Consequently, we do not examine the 8th-grade FKA in this study.

[Table 1 here]

Table 1 shows demographic characteristics of the samples that completed each FKA version. Despite our efforts to recruit a diverse sample, most participants self-identified as non-Hispanic (71% non-Hispanic, 11% Hispanic or Latine, 4% other, 14% non-reported) and White (68% White, 4% Black or African-American, 2% Asian, 8% multiracial, 18% non-reported). Of the 165 families that reported parental education, 92% held a college or university degree, 7% completed partial college or at least one year of specialized training, and 1% listed high school diploma as their most advanced degree.

Materials and General Procedures

The local IRB approved this study. All participants gave oral assent, and their parents or legal guardians signed a consent form. The tasks undertaken by participants are described below.

Fraction Knowledge Assessment (FKA)

We developed the Fraction Knowledge Assessment (FKA), a pencil-and-paper task designed to assess participants' fraction knowledge (for full tests, see our OSF page). Originally developed for adults by Matthews and colleagues (2016), our team adapted the FKA to align with elementary school mathematics curricula. It incorporates items sourced from previous research articles, as well as from international, national, and state achievement exams (including Carpenter, 1981; Hallett et al., 2012, and editions of the NAEP and the TIMSS). Participants completed the FKA in individual sessions held in our laboratory, with no time limit imposed, as part of a larger testing session. On average, participants required approximately 15 minutes to complete the task.

The items included in the FKA versions were aligned with the mathematics school curriculum for 2nd, 3rd, 5th, and 6th grades to avoid ceiling and floor effects. In the first year of the longitudinal study, we used two distinct FKA forms each for 2nd and for 5th grades: Form A was completed by participants from the first enrollment wave, and Form B was completed by those from the second enrollment wave. Younger participants in their final year of the study (i.e., those who began as 2nd graders and concluded as 5th graders) completed FKA 5th Grade Form B. As illustrated in Figure 1, many participants who completed FKA 2nd Grade Form A attained perfect scores, with only a few achieving an accuracy below 50%. Similarly, the distribution of FKA 5th Grade Form A also had a negative skew. To address this bias, Forms B of these FKA versions were developed by maintaining some items from Forms A and introducing others with greater computational complexity.

In all versions of the FKA, participants completed items designed to assess their overall fraction knowledge. Some items remained consistent across FKA versions, serving as anchor items to facilitate equating, whereas others were unique to specific FKA versions. With the exception of 2nd Grade Form A, all FKA versions included items designed to highlight either conceptual (i.e., understanding of fraction properties and fraction magnitude) or procedural fraction knowledge (i.e., fraction operations). The FKA items that highlighted conceptual knowledge investigated abilities related to fraction magnitude understanding, multiplicative reasoning, and fraction density (see full list in OSF). FKA items that highlighted procedural knowledge tested student competence with fraction operations, including addition, subtraction, multiplication, and division. We share the FKA versions via our OSF page, including original task booklets, grade books, and commented FKAs.

When grading the FKAs, each correct answer was awarded one point. Two independent research assistants graded each FKA, following the same rubric (i.e., double-blind grading). In cases where a discrepancy arose between the grades assigned by the two assistants, a third research assistant independently assessed the task without prior knowledge of the initial grades. The final score was determined based on the consensus reached by all three judges. A similar procedure was followed during the curation of the FKA database: two independent research assistants entered item-level data from each FKA into our database. In instances where a discrepancy was noted, a third research assistant entered the data to achieve a consensus. We adopted this approach to minimize the likelihood of errors and to ensure fair scores.

Standardized Measures

In addition to the FKA, students completed the Math Fluency (WJ-MF) and Calculation

(WJ-Calc) subtests of the Woodcock-Johnson III (Woodcock et al., 2001) during each year of their participation in the study in the same session they completed FKA. These tasks are known for having good reliability (>.90) and concurrent validity (correlation with other mathematics tests >.50; Woodcock et al., 2001). The WJ-MF assesses arithmetic fluency by requiring participants to complete 160 simple calculation problems, including addition, subtraction, and multiplication, within a time limit of 3 minutes. One point was given for each correct answer. The WJ-Calc subtest evaluates written calculation skills by presenting participants with 45 problems of increasing complexity, ranging from basic addition to trigonometry. Participants had no time limit to complete these test, and one point was given for each correct answer.

Analyses

We conducted five primary analyses, all implemented in R (R Development Core Team, 2005). First, we examined the reliability and concurrent validity of the FKA using CTT. These analyses were carried out using the "psych" package (Revelle & Revelle, 2015). Second, we conducted a confirmatory factor analysis (CFA) on the FKA versions to assess the fit of the conceptual and procedural fraction knowledge constructs across grades. This analysis was performed using the "mirt" package (Chalmers et al., 2016). Third, we conducted IRT analyses to assess participants' abilities using both a Rasch model and a two-parameter (2PL) model implemented with the "ltm" (Rizopoulos, 2006) and "irtoys" (Partchev et al., 2017) packages. The Rasch model examines how the probability of a correct response in a task is predicted by participants' abilities and items' difficulties, assuming a constant discrimination parameter, whereas the 2PL model adds an item-specific discrimination parameter. We compared the fit of these models to inform our choice of method for further analyses. Fourth, using the "plink" package (Weeks, 2010), we employed Stocking-Lord linking constants to link the ability scores of the different FKA versions, estimated from the 2PL model, based on the anchor items. The anchor items were those that remained identical across FKA 2nd Grade Forms A and B, and 3rd Grade, and across FKA 5th Grade Forms A and B, and 6th Grade. This approach enhances the investigation of development by aligning the different FKAs to a common metric, despite the use of distinct versions of the task across grades. Finally, we examined how fraction knowledge developed by analyzing data from participants that completed the FKA in two consecutive years (i.e., 2nd and 3rd grades for the Younger cohort and 5th and 6th grades for the Older cohort).

Results

Descriptive

As shown in Figure 1, participants did not exhibit floor effects at the group level for any FKA versions. However, some FKA versions displayed a negative skew, indicating that the majority of participants were able to successfully solve the items intended for their respective grades. As the different FKA versions had different numbers of items, here we converted raw accuracy to percent correct to facilitate comparison. Mean accuracy for FKA 2nd Grade Form A was 73% (IQR = [61%, 88%]) and for FKA 2nd Grade Form B was 52% (IQR = [37%, 66%]). As anticipated, our manipulation to increase the difficulty of FKA 2nd Grade from Version A to Version B was effective. The FKA 3rd grade had a mean accuracy of 73% (IQR = [63%,88%]). The mean scores were the same for the FKA 5th Grade Form A (mean accuracy = 66%, IQR = [48%, 80%]) and FKA 5th Grade Form B (mean accuracy = 66%, IQR = [51%, 82%]), suggesting that our difficulty correction was less effective for this FKA version. The FKA 6th grade had a mean accuracy of 73% (IQR = [62%, 86%]). We report further details on the distributions, including mean raw scores, standard-deviation, and IQR, in Supplementary Materials. The differing number of items, along with distinct distributions across the various

FKAs, highlighted the necessity of conducting a psychometric analysis of the task and implementing an equating procedure to compare participants' performance across versions.

[Figure 1 here]

Classical Test Theory Analysis

Cronbach's alpha was high (all alpha > .90) for each FKA version, suggesting high internal consistency across all versions. To assess the concurrent validity of FKAs, we explored their correlation with standardized math tasks (i.e., WJ-MF and WJ-Calc). As indicated in Table 2, all FKAs exhibited moderate to high correlations with the standardized tests, suggesting good concurrent validity across all FKA versions. Overall, the correlations between FKA and WJ-MF and WJ-Calc were comparable to the correlations between WJ-MF and WJ-Calc (see Supplementary Materials).

[Table 2 here]

Factor Structure of FKA

As we designed FKA items to emphasize either conceptual or procedural knowledge, we used confirmatory factor analysis (CFA) to assess the hypothesis that items would load onto these two factors. The FKA 2nd Grade Form A was not included in these analyses since its items were designed to highlight only conceptual fraction knowledge. Contrary to our predictions, results did not provide strong evidence of a definite factor structure for the FKAs. One-dimensional and two-dimensional models had similar fit for FKA 2nd Grade Form B, 3rd Grade, 5th Grade Form A, and 5th Grade Form B. All one-dimensional models had acceptable fits (all RSMAs ≤.08; see Supplementary Materials for details). Moreover, when we explored a two-dimensional solution, the two factors were highly correlated with each other (correlations ranging from .75 to .89), underscoring the plausibility of the one-dimensional solution. The only exception was the 6th Grade FKA.

For 6th Grade FKA, the CFA with the conceptual and procedural factors set *a priori* failed to converge. Thus, we conducted an exploratory factor analysis (EFA). The EFA suggested an alternative two-dimensional solution, diverging from the conceptual and procedural framework. However, the alternative solution performed worse than a one-dimensional model. Furthermore, the factors in the two-dimensional solution were still highly correlated ($r = .69$). These findings indicate that the traditional division between conceptual and procedural fraction knowledge does not present as statistically distinct dimensions with the FKA, consistently with prior research on the factor structure of other fraction and rational number measures (e.g., Hallett et al., 2012; Schneider & Stern, 2010). Since all FKA versions demonstrated an acceptable one-dimensional structure, and factors from two-dimensional models were highly correlated, we assumed unidimensionality for conducting the IRT analyses and linking the FKA versions.

Item Response Theory Analyses

After analyzing the psychometric properties of the FKA versions using CTT and finding evidence that unidimensionality can be assumed, we proceeded with IRT analyses. To investigate FKA items' difficulty and discrimination, we conducted Rasch and 2PL models without including any prior information on individual items. A likelihood ratio test indicated that the Rasch and 2PL models did not fit equally well. Overall, AIC values indicated a better fit for the 2PL model across all FKA versions when compared to the Rasch model. Similarly, log-likelihood analyses favored the 2PL model across all FKA versions compared to the Rasch model. However, BIC values favored the 2PL model only for FKA 3rd Grade and 5th Grade Form B, while indicating better fit for the Rasch model in the other FKA versions. The fit values for both models are shown in Supplementary Materials by FKA version. Given the overall better fit estimates for the 2PL model, and the combination of difficulty and discrimination providing more information for linking the different FKA versions, we decided to use the 2PL model in the following analyses.

We adopted Baker's (2001) proposed cut-off to assess the 2PL indices. Difficulty indices below zero were considered easy, whereas those above zero were considered harder. Discrimination indices greater than 1.35 were considered to indicate high to very high discrimination. Table 3 presents the results of the 2PL model. In general, the FKA versions had more easy than difficult items (mean difficulty index varying from -1.70 to 0.06) but with average to high discrimination (mean discrimination index varying from 1.49 to 1.72).

[Table 3 here]

Items with the lowest and highest difficulty indices tended to have lower discrimination indices across all grades. Overall, item difficulty and discrimination shifted by grade, which may reflect the progress of the mathematics school curriculum. In the FKAs completed by the Younger participants (i.e., 2nd and 3rd graders), items with higher difficulty and discrimination assessed magnitude understanding. In contrast, in the FKAs completed by Older participants (i.e., 5th and 6th graders), the items with higher difficulty and discrimination required solving operations, especially division, with mixed numbers and fractions without a common denominator. We summarize the difficulty and discrimination metrics of each FKA version in Supplementary Materials, and give item-by-item details on the OSF page.

Linking FKA Versions

To link the different FKA versions, we used FKA 2nd Grade Form B and 5th Grade Form B as references, due to their larger number of shared items with the other FKA versions. Using FKA 2nd Grade Form B as a reference, we linked the ability scores of FKA 2nd Grade Form A

(30 overlapping items) and FKA 3rd Grade (20 overlapping items) to it. Similarly, we linked the FKA 5th Grade Form B with those of FKA 5th Grade Form A (33 overlapping items) and FKA 6th Grade (13 overlapping items). Stocking-Lord linking constants were employed for this process, which involved calculating two linear transformation constants— a slope A and an intercept B—to standardize a set of items against a target (Stocking & Lord, 1983). Reference ability scores and linked ability scores (logits) are described in Table 4.

[Table 4 here]

Fraction Knowledge Development

After confirming evidence for the validity and reliability of FKA and equating scores across its versions, we proceeded with developmental analyses. We explored the impact of an additional year of schooling on students' fraction knowledge. Given that formal fraction instruction typically begins in the 3rd grade, we anticipated larger improvements in participants' fraction knowledge from 2nd to 3rd grade than from 5th to 6th grade. While we predicted overall improvements in fraction knowledge, prior longitudinal studies have suggested varied developmental trajectories (Ye et al., 2016), raising the possibility that some participants would exhibit similar ability levels across grades or even decreases in the second assessment. Contrasting participants' FKA scores between grades allows us not only to observe developmental trends in fraction knowledge but also to assess how FKA detects these changes.

We analyzed participants' fraction knowledge across grades using the linked ability scores. For these analyses, we only included participants who completed FKA in both 2nd and 3rd grades or both 5th and 6th grades. Thus, the sample with 2nd and 3rd graders was composed of 142 participants, and the sample with 5th and 6th graders was composed of 113 participants. Results are illustrated in Figure 2. Out of the 142 participants who completed FKA in 2nd and 3rd grades, 135 (95%) improved their scores with grade. In 2^{nd} grade, participants' mean linked ability score was 0.27 (*sd* = 1.23, *min* = -3.53, *max* = 4.31), whereas in 3^{rd} grade, their mean linked ability score was 1.76 (*sd* = 1.20, *min* = -1.41, *max* = 5.85). This difference corresponds to an improvement in their fraction knowledge of about one and a half standard deviations of the ability scores, $t(141) = 18.08$, $p < .001$, $d = 1.22$, indicating a higher average level of the latent trait being measured when students were in 3^{rd} grade compared to when they were in 2^{nd} grade. Some improvements were also observed from 5^{th} to 6^{th} grade, but they were less robust. Out of the 113 participants who completed FKA in both grades, 78 (69%) had higher scores in 6^{th} grade. Participants' mean linked ability scores were -0.02 in 5^{th} grade (*sd* = 1.19, *min* = -2.75, *max* = 3.11) and 0.27 in 6^{th} grade (*sd* = 1.02, *min* = -2.09, *max* = 3.56), which was a significant improvement, despite being weaker than the one observed from 2^{nd} to 3^{rd} grade, $t(112) = 3.50$, $p = .001$, $d = 0.26$. These results demonstrate stronger developmental shifts in fraction knowledge in the transition from 2^{nd} to 3^{rd} grade, when students are first introduced to fractions in school.

We also examined how participants' fraction knowledge in an earlier grade influenced their fraction knowledge in a subsequent grade. Better FKA linked ability scores in 2^{nd} grade were associated with better linked ability scores in 3^{rd} grade, $r(140) = .68$, $p<.001$. Similar results were found for 5^{th} and 6^{th} grades, $r(111) = .69$, $p<.001$ (Figure 2). These findings suggest that participants with stronger fraction knowledge in an earlier grade were more likely to achieve better FKA scores in the subsequent grade.

[Figure 2 here]

Discussion

The finding that fraction knowledge plays a critical role in students' transition from basic to high-order mathematics (e.g., Booth et al., 2012) has increased interest in understanding the developmental trajectory, predictors, and outcomes of fraction knowledge (Siegler et al., 2013). However, many studies have not reported the psychometric properties of their measures, have failed to adapt measures across different grade levels, and have paid limited attention to fraction knowledge in early grades (e.g., Bailey et al., 2014; Fazio et al., 2016; Hansen et al., 2015; Kalra et al., 2020; Matthews et al., 2016; Ye et al., 2016), making comparing and generalizing findings across previous studies challenging. To address these gaps, we investigated fraction knowledge in 2nd, 3rd, 5th and 6th grades using grade-appropriate measures with high face-validity, designed and psychometrically tested by our research team: the FKA.

Our results provide evidence for the reliability and validity of the FKA as a tool for studying the development of fraction knowledge across the elementary school years. Classical test theory demonstrated that all FKAs had good internal consistency and concurrent validity, and factor analyses indicated that FKA has a one-dimensional structure. IRT analyses indicated a good balance between difficulty and discrimination indices. In addition to providing evidence for robust psychometric properties of FKA, our study also contributes to conceptualizing the factor structure of the fraction knowledge construct, supports fraction knowledge assessment from earlier grades, and may benefit educators, psychologists, and researchers interested in assessing fraction knowledge, since we are freely sharing the FKAs.

The Factor Structure of Fraction Knowledge

Overall, our results contribute to the ongoing debate regarding the appropriateness of characterizing fractions knowledge as separable into distinct procedural and conceptual components. We drew from prior research that has conceptualized fraction knowledge as consisting of both conceptual and procedural factors (Booth & Newton, 2012; Lenz et al., 2020; Siegler et al., 2013), and therefore designed the FKA to include items targeting both conceptual and procedural fraction knowledge (except for the 2nd-Grade A version). However, when we evaluated this factor structure using CFA, our results did not clearly support a two-dimensional model with conceptual and procedural factors. Instead, both one-dimensional and two-dimensional models demonstrated similar fit indices. Moreover, in the two-dimensional models, the conceptual and procedural factors were highly correlated, suggesting considerable overlap between the two constructs. Subsequent EFA suggested a reasonable fit for a two-dimensional structure that differed from the conceptual and procedural factors. However, even then the one-dimensional model still showed good fit and the new factor solution had highly correlated factors. These findings align with previous research that has also failed to establish a clear factor structure in fraction knowledge tasks (Schneider & Stern, 2010) or found that two-dimensional solutions had highly correlated factors (Hecht et al., 2003; Hallet et al., 2010).

There is conflicting evidence regarding the distinction between conceptual and procedural fraction knowledge. While some studies have found that these factors are relatively independent (e.g., Lenz et al., 2020), other studies have failed to find a separation between them (e.g., Schneider & Stern, 2010). These inconsistencies may stem from variations in how these constructs are defined and operationalized (Alibali & Sidney, 2015; Crooks & Alibali, 2014; Rittle-Johnson et al., 2015). However, we defined conceptual and procedural knowledge similarly to previous studies that have identified separate factors. We operationalized conceptual fraction knowledge as fraction magnitude understanding (similar to Hecht et al., 2003; Lenz et al., 2020), and procedural fraction knowledge as the ability to perform operations with fractions, aligned with those used in previous studies that have successfully identified a factor structure in fraction knowledge assessments (e.g., Hecht et al., 2003; Lenz et al., 2020). Thus, we do not think that different operationalizations can explain the discrepancy between our results and previous studies that do identify separate conceptual and procedural factors.

Another explanation for the conflicting findings regarding the separation of conceptual and procedural fraction knowledge may stem from diversity in participants' strategy application. The diversity in strategies employed across students to solve the same problems may depend on the problem's characteristics, students' prior knowledge, and contextual factors (Alibali & Sidney, 2015). For example, items designed to measure conceptual knowledge may inadvertently tap into procedural knowledge. For instance, Matthews and Rittle-Johnson (2009) demonstrated that some 2^{nd}- to 5^{th}-grade children solve whole number problems intended to assess conceptual mathematics knowledge (e.g., "3+2 = 7-2") by employing procedural strategies (e.g., solving 3+2=5 first, then 7-2=5, and finally checking for equality). Similarly, we suggest that some of our participants may have used conceptual knowledge to solve some FKA items designed to measure procedural knowledge, and vice versa. For instance, participants could address an item focused on conceptual knowledge, such as "how many fourths do you need to make a whole?" by applying the procedural method of adding fourths until reaching a whole and then counting the number of operands. If participants employed a mixture of conceptual and procedural knowledge to solve items intended to target only one of these constructs, a clear factor structure might not emerge, leading to a one-dimensional model with good fit, as we found for FKA.

The challenges associated with distinguishing between conceptual and procedural fraction knowledge may not be unique to our measure, but rather inherent to the fraction knowledge construct itself (Rittle-Johnson & Schneider, 2015). Instead of being classified into conceptual and procedural domains, fraction knowledge may have a one-dimensional structure, as our findings suggest. To gain a deeper understanding of the underlying structure of fraction knowledge, it is imperative to explore alternative latent factors using data-driven approaches, such as EFA. Here, we conducted EFA only on the 6th grade version of the FKA; future investigations should examine the factor structure of all the versions of the FKA. Moreover, it is crucial to examine how the structure of fraction knowledge changes across development, as children may adjust, discard, or adopt new problem-solving strategies over time (Siegler, 1976), potentially giving rise to emergent factors. In addition to data-driven approaches, future work should refine our conceptual frameworks, such as the distinction between conceptual and procedural fraction knowledge. Developing a robust theory about the fraction knowledge constructs will enable the creation of measures aligned with them, which can then be rigorously evaluated using data-driven approaches.

Assessing early fraction knowledge with the FKA

Most fraction knowledge measures have primarily focused on more advanced grades, leaving a gap in our understanding of early fraction skills (e.g., Jordan et al., 2017; McMullen et al., 2015; but see Viegut et al., 2023). In this study, one of our aims was to support early fraction knowledge assessment by developing an appropriate measure targeting 2nd graders. Our findings revealed that the FKA is indeed appropriate to 2nd graders, as both 2nd-Grade A and B FKA versions had good difficulty and discrimination indices. Notably, 2nd graders had high proficiency in FKA items that assessed nonsymbolic ratio comparison and estimation, and the mapping of nonsymbolic and symbolic representations of halves and fourths. These results indicate that fraction knowledge, especially the ability to process different representations of "half", may start developing even before formal introduction to fractions in schools, supporting findings from previous research (Barth et al., 2009; Mix, 1999; Park et al., 2021; Spinillo & Bryant, 1991; Viegut et al., 2023). These results highlight the importance of assessing fraction knowledge in early grades, which can be achieved with FKA.

FKA is an appropriate tool to investigate fraction knowledge across development

One important contribution of the current study is to show how linking grade-appropriate fraction knowledge tasks allow for comparing across grade levels using the same metric. Importantly, the ability levels allow us to predict how a student might perform on a problem of a given difficulty level even if they have not faced the problem yet. In particular, by linking different FKA versions, we were able to use this measure to track changes in students' fraction knowledge with one additional year of schooling. Our longitudinal analyses indicated significant improvements in participants' fraction knowledge from 2nd to 3rd grade. In contrast, the progress from 5th to 6th grade was more modest. This trend also aligns with the mathematics curriculum, which dedicates a substantial portion of direct instruction to fractions in 3rd grade but emphasizes it less by 6th grade (National Governors Association, 2010). These findings underscore the critical importance of early formal fraction instruction, suggesting that the initial years of schooling represent a pivotal window for fraction knowledge development. Given that prior fraction knowledge strongly predicts future high-order mathematics skills (Booth et al., 2012; Hansen et al., 2017), effective fraction assessment in the early grades holds particular significance, especially for children at risk for mathematics difficulties.

Our findings indicate that the FKA is effective in identifying participants' main fraction difficulties and discriminating their abilities across different grades. Notably, the difficulty and discrimination indices of the FKA items varied across grade levels, suggesting that a common fraction knowledge assessment may not adequately capture students' progress over time. In 2nd and 3rd grades, items assessing fraction magnitude understanding had higher discrimination and difficulty levels. Conversely, these items were among the easiest in more advanced grades. In 5th and 6th grades, items with higher discrimination and difficulty focused on operations, particularly division, involving mixed numbers and fractions without common denominators. The grade-related variations in item discrimination and difficulty for similar skills may reflect shifts in participants' fraction knowledge, likely related to their progression through the mathematics curriculum: the Mathematics Common Core Standards emphasize fraction magnitude understanding in earlier grades and fraction operations in later grades (National Governors Association, 2010). This alignment of FKA difficulty and discrimination with the curriculum suggests this task can effectively capture students' fraction knowledge growth across grades.

Applications of the FKA

Fraction knowledge is recognized as foundational for higher-order mathematics skills (Booth & Newton, 2012). Consequently, employing reliable and valid measures to assess fraction knowledge is imperative for identifying children who may be at risk and for evaluating the effectiveness of interventions. Currently available fraction knowledge measures often lack comprehensive coverage across different grades and detailed reporting of their psychometric properties (e.g., Rodrigues et al., 2019; Viegut et al., 2023). In this study, we share the FKAs free of cost via OSF (along with rubric, commented FKA versions, and with scripts to run the linking procedure), allowing them to be used either in their original form or adapted for specific needs. With demonstrated evidence of validity and reliability, the FKA offers a balance between students' abilities and task difficulty, ensuring an unbiased assessment. Importantly, when used with linking procedures, the FKA can track the developmental trajectory of fraction knowledge. In the future, the FKA should be adapted to keep the best items, making an efficient version that allows for quick assessments that can be used to place students on an ability map scaled vertically with this interval-level metric. Further, items could be included in adaptive computer-based tasks. Given its potential significance and versatility, the FKA stands as a valuable tool for researchers, educators, curriculum developers, and policymakers, allowing the assessment of fraction knowledge both concurrently and longitudinally to draw well-informed decisions regarding diagnosis, interventions, and academic placements.

Limitations and Future Directions

This study is not without its limitations. Currently, the FKA enables the investigation of fraction knowledge only in 2nd, 3rd, 5th, and 6th grades. Although we initially aimed to include 8th-grade FKA in this analysis, COVID-19 disruptions prevented data collection when participants reached this grade. Moving forward, future investigations should expand the FKA to additional grade levels and use the linking methods outlined in this study to comprehensively track fraction knowledge development throughout elementary school. Using the linking procedure, it is possible to use the FKA for comprehensive developmental studies, including but not limited to latent growth, cross-lagged panel, and structural equation models, which should be targets of future investigations. Future research could leverage new FKA versions and methods of analyses to track developmental trajectories. Additionally, as the FKA has some open-ended items, it allows for a qualitative analysis of students' errors—which may inform us about their strategies and the type of knowledge they engage in to solve distinct fraction problems across developmental phases.

Moreover, it is imperative for future research to extend the application of FKA to different populations and educational settings. Despite efforts to recruit a representative sample, our participants were predominantly non-Hispanic, white, and from middle-class backgrounds. To ensure the generalization of our results regarding the FKA, it is essential to adapt it, to examine its properties, and to establish norms with diverse samples, including students from various socioeconomic and linguistic backgrounds. It is also important to adapt the FKA to other educational contexts. Most of the data collection for this study occurred before the onset of the COVID-19 pandemic. Considering the significant disruptions caused by the pandemic in the educational landscape, with delays in learning reaching up to two years (Donnelly & Patrinos, 2021; Starling-Alves et al., 2023), it is crucial to evaluate the properties of the FKA among cohorts of students who experienced school closures. This will provide valuable insights into how the pandemic may have impacted fraction knowledge development and the efficacy of educational interventions. Finally, this study does not explore the main predictors of fraction knowledge development and how fraction knowledge explains high-order mathematics across grades. Moving forward, future research should use the FKA to track the developmental shifts in fraction knowledge while examining its interplay with other cognitive abilities and advanced mathematics skills. We are making the FKA available as a free tool to encourage the academic community to use it in subsequent studies. This collaborative effort will contribute to a deeper understanding of fraction knowledge development and its implications for mathematics education.

Conclusion

Reliable and valid tools are crucial to effectively track children's fraction knowledge development via formative and summative classroom-based assessments and may help identifying children at risk for math difficulties. In this study, we presented a novel fraction knowledge measure—the FKA. Results indicated that the FKA is a one-dimensional task with good internal consistency and concurrent validity. Importantly, the one-dimensional structure of FKA may reflect an underlying one-dimensional structure of the fraction knowledge construct. A longitudinal analysis with linked FKA scores showed that fraction knowledge development was more pronounced from 2nd to 3rd grade than from 5th to 6th grade. We share the FKA freely via

OSF, so it can be used by researchers, educators, curriculum developers, and policymakers to promote a better understanding of children's developing fraction knowledge.

References

Alibali, M. W., & Sidney, P. G. (2015). Variability in the natural number bias: Who, when, how, and why. *Learning and Instruction, 37,* 56-61. doi: 10.1016/j.learninstruc.2015.01.003

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*(3), 447-455. doi: 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*(5), 775-785. doi: 10.1111/desc.12155.

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 US and Chinese children. *Journal of Experimental Child Psychology, 129*, 68-83. doi: 10.1016/j.jecp.2014.08.006

Baker, F. B. (2001). *The basics of item response theory.* Available at: http://ericae.net/irt/baker

Barth, H., Baron, A., Spelke, E., & Carey, S. (2009). Children's multiplicative transformations of discrete and continuous quantities. *Journal of Experimental Child Psychology*, *103*(4), 441-454. doi: 10.1016/j.jecp.2009.01.014

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

Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Lindquist, M. M., & Reys, R. E. (1981). What are the chances of your students knowing probability?. *The Mathematics Teacher*, *74*(5), 342-344.

Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., KwonHyun, K., Falk, C. F., & Meade, A. (2015). Package 'mirt'. Zugriff AM, 21(01).

Chan, W. H., Leu, Y. C., & Chen, C. M. (2007). Exploring group-wise conceptual deficiencies of fractions for fifth and sixth graders in Taiwan. *The Journal of Experimental Education, 76*(1), 26-57. doi: 10.3200/JEXE.76.1.26-58

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental review, 34*(4), 344-377. doi: 10.1016/j.dr.2014.10.001

Donnelly, R., & Patrinos, H. A. (2021). Learning loss during Covid-19: An early systematic review. *Prospects, 51,* 601–609. doi: 10.1007/s11125-021-09582-6

Dorri, M. M., & Rafiepour, A. (2018). Evaluating of Rational Number Misconception Test (RNMT) with Rasch model. *Quarterly of Educational Measurement, 8*(32), 103-126. doi: 10.22054/jem.2019.31422.1728

Eason, S. H., & Ramani, G. B. (2020). Parent–child math talk about fractions during formal learning and guided play activities. *Child Development, 91*(2), 546-562. doi: 10.1111/cdev.13199

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

Fazio, L. K., Kennedy, C. A., & Siegler, R. S. (2016). Improving children's knowledge of fraction magnitudes. *PloS one, 11*(10), e0165243. doi: 10.1371/journal.pone.0165243

Gelman, R. (1991). Epigenetic foundations of knowledge structures: Initial and transcendent constructions. *The epigenesis of mind: Essays on biology and cognition*, 293-322.

Hallett, D., Nunes, T., Bryant, P., & Thorpe, C. M. (2012). Individual differences in conceptual and procedural fraction understanding: The role of abilities and school experience. *Journal of Experimental Child Psychology, 113*(4), 469-486. doi: 10.1016/j.jecp.2012.07.009

Hansen, N., Jordan, N. C., Fernandez, E., Siegler, R. S., Fuchs, L., Gersten, R., & Micklos, D. (2015). General and math-specific predictors of sixth-graders' knowledge of fractions. *Cognitive Development*, *35*, 34-49. doi: 10.1016/j.cogdev.2015.02.001

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

Hecht, S. A., & Vagi, K. J. (2012). Patterns of strengths and weaknesses in children's knowledge about fractions. *Journal of Experimental Child Psychology*, *111*(2), 212-229. doi: 10.1016/j.jecp.2011.08.012

Hurst, M. A., & Cordes, S. (2018). Attending to relations: Proportional reasoning in 3-to 6-year-old children. *Developmental Psychology, 54*(3), 428. doi: 10.1037/dev0000440

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. doi: 10.1016/j.jecp.2013.02.001

Jordan, N. C., Resnick, I., Rodrigues, J., Hansen, N., & Dyson, N. (2017). Delaware longitudinal study of fraction learning: Implications for helping children with mathematics difficulties. *Journal of Learning Disabilities*, *50*(6), 621-630. doi: 10.1177/0022219416662033

Kalra, P. B., Hubbard, E. M., & Matthews, P. G. (2020). Taking the relational structure of fractions seriously: Relational reasoning predicts fraction knowledge in elementary school children. *Contemporary Educational Psychology*, *62*, 101896. doi:

10.1016/j.cedpsych.2020.101896

Lee, W. C., & Ban, J. C. (2009). A comparison of IRT linking procedures. *Applied Measurement in Education*, *23*(1), 23-48. doi: 10.1080/08957340903423537

Lee, V. R. (2011). The relationship between special education teachers' mathematical knowledge, instructional choices, and beliefs about teaching and learning mathematics. State University of New York at Albany.

Lenz, K., Dreher, A., Holzäpfel, L., & Wittmann, G. (2020). Are conceptual knowledge and procedural knowledge empirically separable? The case of fractions. *British Journal of Educational Psychology, 90*(3), 809-829. doi: 10.1111/bjep.12333

Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment*, *1*(1), 1-11.

Matthews, P. G., Lewis, M. R., & Hubbard, E. M. (2016). Individual differences in nonsymbolic ratio processing predict symbolic math performance. *Psychological Science, 27*(2), 191-202. doi: 10.1177/0956797615617799

McMullen, J., Laakkonen, E., Hannula-Sormunen, M., & Lehtinen, E. (2015). Modeling the developmental trajectories of rational number concept (s). *Learning and Instruction*, *37*, 14-20. doi: 10.1016/j.learninstruc.2013.12.004

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology*, *35*(1), 164. doi: 10.1016/B978-0-12-805086-6.00006-0

National Center for Education Statistics. (2017). NAEP Mathematics Report Card: National Achievement-Level Results. Available at: https://www.nationsreportcard.gov/math_2017/nation/achievement?grade=8

National Governors Association. (2010). *Common core state standards*. Washington, DC.

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

Neagoy, M. (2017). *Unpacking fractions: Classroom-tested strategies to build students' mathematical understanding*. ASCD.

Ölmez, İ. B., & Izsák, A. (2021). Characterizing reasoning about fraction arithmetic of middle grades teachers in three latent classes. *Mathematical Thinking and Learning, 23*(3), 225-253. doi: 10.1080/10986065.2020.1780368

Park, Y., Viegut, A. A., & Matthews, P. G. (2021). More than the sum of its parts: Exploring the development of ratio magnitude versus simple magnitude perception. *Developmental Science*, *24*(3), e13043.

Partchev, I., Partchev, M. I., & Suggests, M. A. S. S. (2017). *Package 'irtoys'. A collection of functions related to item response theory (IRT).*

R Development Core Team. (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Revelle, W., & Revelle, M. W. (2015). Package 'psych'. *The comprehensive R archive network*, 337, 338.

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. *Oxford Handbook of Numerical Cognition*, 1118-1134. doi: 10.1093/oxfordhb/9780199642342.013.014

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346. doi: 10.1037/0022-0663.93.2.346

Rizopoulos, D., & Rizopoulos, M. D. (2018). *Package 'ltm'*. Available at: http://wiki. r-project.org/rwiki/doku.php.

Rodrigues, J., Jordan, N. C., & Hansen, N. (2019). Identifying fraction measures as screeners of mathematics risk status. *Journal of Learning Disabilities*, *52*(6), 480-497. doi: 10.1177/0022219419879684

Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental Psychology, 46*(1), 178. doi: 10.1037/a0016701

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*(4), 481-520. doi: 10.1016/0010-0285(76)90016-5

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

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: The new frontier for theories of numerical development. *Trends in Cognitive Sciences, 17*(1), 13-19. doi: 10.1016/j.tics.2012.11.004

Spinillo, A. G., & Bryant, P. (1991). Children's proportional judgments: The importance of "half". *Child Development*, *62*(3), 427-440. doi: 10.2307/1131121

Starling-Alves, I., Hirata, G., & Oliveira, J. B. A. (2023). Covid-19 school closures negatively impacted elementary-school students' reading comprehension and reading fluency skills. *International Journal of Educational Development, 99*, 102753. doi: 10.1016/j.ijedudev.2023.102753

Stigler, J. W., Givvin, K. B., & Thompson, B. J. (2010). What community college developmental mathematics students understand about mathematics. *MathAMATYC Educator, 1*(3), 4-16.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*(2), 201-210. doi: 10.1177/014662168300700208

Van Hoof, J., Janssen, R., Verschaffel, L., & Van Dooren, W. (2015). Inhibiting natural knowledge in fourth graders: towards a comprehensive test instrument. *ZDM, 47*, 849-857. doi: 10.1007/s11858-014-0650-7

Viegut, A. A., Resnick, I., Miller-Cotto, D., Newcombe, N., & Jordan, N. (2023). Tracking Informal Fraction Knowledge and Its Correlates Across First Grade.

Wang, Y., & Siegler, R. S. (2023). Developmental trajectories of numerical magnitude representations of fractions and decimals. *Developmental Psychology*, *59*(5), 874-885. doi: 10.1037/dev0001547

Weeks, J. P. (2010). plink: An R package for linking mixed-format tests using IRT-based methods. *Journal of Statistical Software, 35*, 1-33. doi: 10.18637/jss.v035.i12

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Examiner's Manual, Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing Company.

Ye, A., Resnick, I., Hansen, N., Rodrigues, J., Rinne, L., & Jordan, N. C. (2016). Pathways to fraction learning: Numerical abilities mediate the relation between early cognitive competencies and later fraction knowledge. *Journal of Experimental Child Psychology, 152*, 242-263. doi: 10.1016/j.jecp.2016.08.001

Tables and Figures

Table 1. Demographic characteristics

FKA Version	**N**	**Age**		**Gender**		
		Mean	***SD***	***Male***	***Female***	***Non-reported***
2nd Grade Form A	89	8.11	0.64	48	33	8
2nd Grade Form B	97	8.01	0.56	55	40	2
3rd Grade	163	9.00	0.87	98	61	4
5th Grade Form A	54	10.99	0.58	25	26	3
5th Grade Form B	118	10.94	0.55	65	51	2
6th Grade	114	11.95	0.53	63	47	4

Table 2. FKA internal consistency and concurrent validity

FKA Version	Cronbach's alpha	WJ-MF (r)	WJ-Calc (r)
2nd Grade Form A	.90	.43***	.57***
2nd Grade Form B	.90	.50***	.68***
3rd Grade	.92	.41***	.54***
5th Grade Form A	.93	.55***	.61***
5th Grade Form B	.91	.51***	.67***
6th Grade	.91	.43***	.69***

Note. *** = $p<.001$

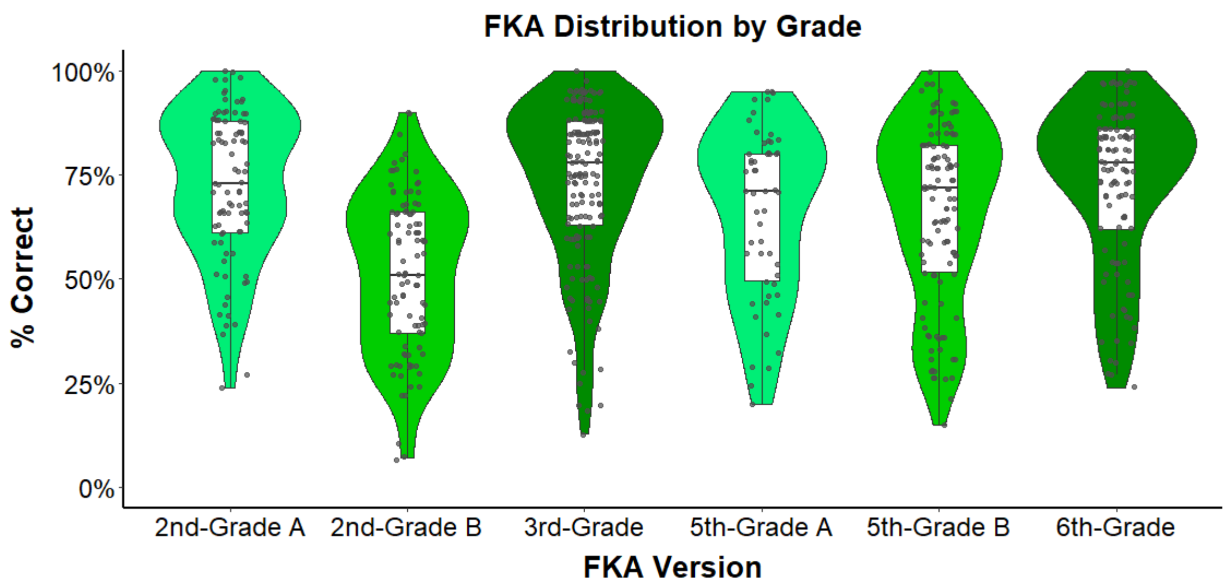
Table 3. FKA versions' mean difficulty and discrimination indices

FKA Version	Difficulty Index		Discrimination Index	
	Mean	*SD*	*Mean*	*SD*
2nd Grade Form A	-1.70	1.96	1.54	0.79
2nd Grade Form B	0.06	1.91	1.49	0.79
3rd Grade	-1.15	1.29	1.72	0.83
5th Grade Form A	-0.99	1.75	1.57	0.79
5th Grade Form B	-1.04	1.45	1.61	0.89
6th Grade	-1.05	1.57	1.49	0.74

Table 4. Participants' Ability Scores Estimated for Each FKA Version

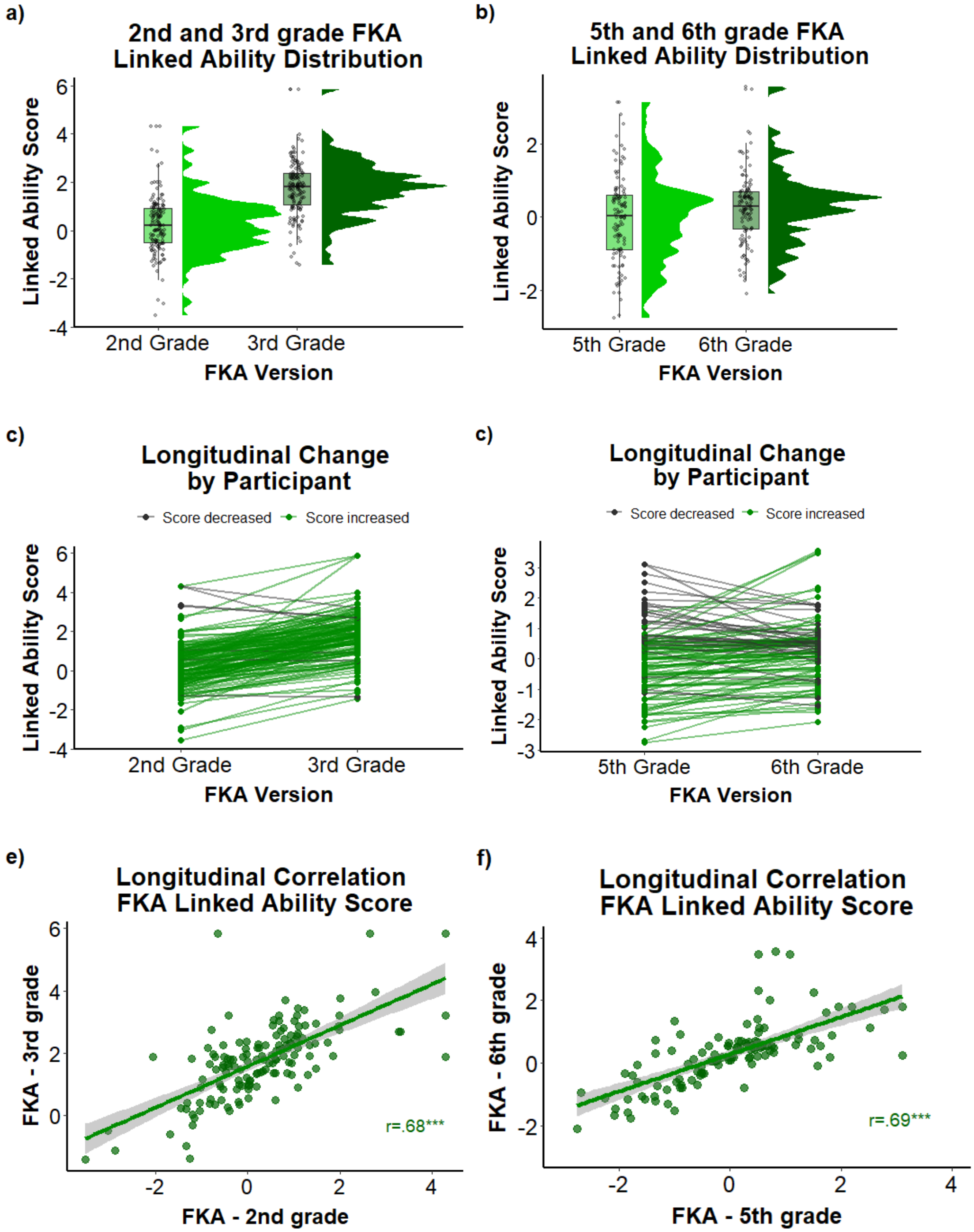
Cohort	FKA Version	Number of Linking Items	θ	
			Mean	*SD*
Younger	**2nd Grade Form A – linked ability score**	30	0.40	1.23
	2nd Grade Form B – ability score (reference)	39	0.08	1.13
	3rd Grade – linked ability score	20	1.75	1.20
Older	**5th Grade Form A – linked ability score**	33	0.03	1.14
	5th Grade Form B – ability score (reference)	39	-0.09	1.13
	6th Grade – linked ability score	13	0.27	1.02

Figure 1. Participants' performance (% correct) on the different FKA versions



Note. Violin plots with box-plots representing the distribution of raw scores each FKA version. Each dot represents the score of one participant. No floor effect at the group level was observed. Overall, distributions were left-skewed. FKA versions 2nd-Grade A and 6th-Grade had the highest mean scores, and 2nd-Grade B had the lowest mean score. In the plot, the y-axis was converted to percentage to facilitate comparison across FKA versions.

Figure 2. Fraction Knowledge Across Grades



Note. Plots on the left side represent data from younger participants and plots on the right side represent data from older participants. Plots show changes in FKA linked ability scores distribution from 2nd to 3rd grade (Figure 2a) and from 5th to 6th grade (Figure 2b), individual changes in FKA linked ability scores from 2nd to 3rd grade (Figure 2c) and from 5th to 6th grade (Figure 2d), and the correlation between 2nd and 3rd grade FKA linked ability scores (Figure 2e), and 5th and 6th grade FKA linked ability scores (Figure 2f).