Article:

Di Lonardo Burr, S.M., Xu, C., Li, H. et al. (2026) Finding their way: Students' fraction misconceptions and errors over two years of fraction learning. Learning and Instruction, 106. 102454. ISSN: 0959-4752

https://doi.org/10.1016/j.learninstruc.2026.102454

FINDING THEIR WAY: STUDENTS' FRACTION MISCONCEPTIONS AND ERRORS OVER TWO YEARS OF FRACTION LEARNING

Sabrina M. Di Lonardo Burr[a*], Chang Xu[b], Hongxia Li[c], Jiwei Si[c], Bijuan Huang[d], & Kaiyue Guo[c]

[a] School of Psychology, University of Sheffield, United Kingdom

[b] School of Psychology, Queen's University Belfast, United Kingdom

[c] Shandong Provincial Key Laboratory of Brain Science and Mental Health, Faculty of Psychology, Shandong Normal University, China

[d] School of Psychology, University of Jinan, China

*Corresponding Author: Sabrina M. Di Lonardo Burr, School of Psychology

The University of Sheffield, 219 Portobello, Broomhall, Sheffield, S1 4DP, United Kingdom

(email: s.burr@sheffield.ac.uk)

Sabrina Di Lonardo Burr: https://orcid.org/0000-0001-6338-9621

Chang Xu: https://orcid.org/0000-0002-6702-3958

Hongxia Li: https://orcid.org/0000-0002-5052-2742

Jiwei Si: https://orcid.org/0000-0003-1342-7348

Bijuan Huang: https://orcid.org/0000-0003-0873-9670

Kaiyue Guo: https://orcid.org/0000-0003-4234-3450

This manuscript was accepted for publication in the *Learning & Instruction* on August 3, 2026. This preprint is the peer-reviewed accepted version but has not yet been copyedited and may differ from the final version published in the journal.

Abstract

Purpose. This study examined the longitudinal stability and consistency of student misconceptions across five fraction subconstructs during two years of mathematics instruction.

Method. In this study, 1,055 students (546 boys; M_{age} = 9.8 years) were assessed at four time points in Grades 4 and 5 across five subconstructs: part-whole, measurement, quotient, operator, and density. Using a methodical pre-registered error-classification framework, supplemented by exploratory mixed-effects modelling, we examined the frequency and stability of student errors.

Results. Results revealed a clear distinction in conceptual stability across domains. Part-whole meanings were mastered early in instruction and remained stable over time, whereas tasks involving magnitude and density were characterized by high response entropy and frequent strategy shifts. Students simultaneously held multiple misconceptions, which were highly dependent on the specific subconstruct being assessed. For example, in fraction arithmetic, errors varied across representations and operations. While accuracy generally improved, the progression of errors was often variable, with students frequently reverting from correct responses to whole number bias or unique, unclassified strategies when new topics, such as decimals, were introduced.

Conclusion. These patterns indicate that a correct response at a single time point is not always a sign of stable understanding. Instead, the high degree of variation suggests that early fraction knowledge is context-dependent and that unique error patterns often signal active conceptual reorganization. These findings suggest that because misconceptions are a moving target, interventions must be adaptive and responsive to the specific errors that emerge as students transition between new fraction concepts.

Keywords: fractions; misconceptions; errors; mathematical development

Finding their Way: Students' Fraction Misconceptions and Errors Over Two Years of Fraction Learning

1. Introduction

Difficulties with fraction learning often emerge in elementary school when fractions are first introduced and persist into secondary school and beyond (Siegler & Lortie-Forgues, 2017; Siegler & Pyke, 2013). When students are learning fractions, difficulties are thought to arise from the incorrect application of whole number thinking to rational numbers (Ni & Zhou, 2005; Siegler & Lortie-Forgues, 2017) which can lead to persistent errors and misconceptions (Vamvakoussi et al., 2012). Misconceptions, which reflect fundamental misunderstanding that systematically results in error (Smith et al., 1994), can impede the integration of prior and new knowledge (González-Forte et al., 2023). Unfortunately, identifying misconceptions in the classroom is non-trivial, with (pre)service teachers' knowledge of fraction concepts largely mirroring students' knowledge (Van Steenbrugge et al., 2014). Critically, understanding fraction learning requires examining not only whether students improve, but how their misconceptions change over time. Rather than progressing in a steady manner, students may show periods of regression, instability, and strategy shifting as instruction unfolds. In the present study we explore misconceptions, errors, and their patterns of change over the course of two years of fraction learning, moving beyond accuracy to examine stability, instability, and dynamic changes in student thinking.

1.1 The Multifaceted Nature of Fractions

Part of the challenge of learning fractions is their multifaceted nature (Brousseau et al., 2004). Students need to comprehend interrelated subconstructs of fractions, such as part-whole, ratio, operator, quotient, and measure to demonstrate fraction understanding (Charalambous &

Pitta-Pantazi, 2007; Kieren, 1976). These interrelated subconstructs are linked to fraction operations, equivalence, and problem solving in both children and adults (Behr et al., 1993; Charalambous & Pitta-Pantazi, 2007). Because subconstructs are introduced at different times in the curriculum, misconceptions may emerge and recede at different rates, potentially producing distinct developmental trajectories across tasks and concepts, leading to divergences in student performance across tasks.

A primary source of difficulty is the interference of students' existing whole-number knowledge. When introduced to fractions, students often attempt to fit them into their whole-number framework, processing the whole number components of fractions discretely rather than processing fractions as a single magnitude (Ni & Zhou, 2005). This whole number bias is evident across many tasks. For example, in a basic fraction notation task (i.e., mapping a rectangle with 3 out of 7 equally partitioned sections shaded to the fraction $\frac{3}{7}$), students may only include the numerator in their response (e.g., 3 instead of $\frac{3}{7}$; Corwin et al., 1991; Di Lonardo Burr et al., 2022; Mack, 1995). In fraction magnitude tasks like comparison and ordering, students may directly compare either the numerators or denominators to determine the larger value (Bonato et al., 2007; DeWolf & Vosniadou, 2015; Meert et al., 2010; Malone & Fuchs, 2017) or assume that fractions with larger components are greater than equivalent fractions with smaller components (e.g., $\frac{3}{15} > \frac{2}{10}$; Braithwaite & Siegler, 2018; Ni, 2001). Similarly, when performing operations, students may continue to apply whole number arithmetic strategies, such as adding or subtracting both the numerators and denominators (Braithwaite et al., 2018).

Beyond whole number bias, students may also exhibit misconceptions related to the part-whole concept. For instance, they may understand that fractions consist of two parts but fail to recognize that the parts must be equal in size (Lamon, 2020; Marshall, 1993) or that fractions do not refer to discrete countable pieces (Saxe et al., 2005). For example, returning to our rectangle with 3 out of 7 equally partitioned sections shaded, students might label this fraction as "$\frac{3}{4}$", erroneously believing that the numerator and denominator represent the shaded and unshaded partitioned sections, respectively (Di Lonardo Burr et al., 2022). With respect to fraction magnitude, students may develop a partial understanding in which they recognize that proper fractions represent quantities smaller than whole numbers and that fraction magnitude does not follow whole-number logic (Stafylidou & Vosniadou, 2004). However, they may still struggle to coordinate the numerator and denominator when comparing fractions, sometimes relying on overgeneralized intuitions about part whole relations (e.g., believing that fractions with smaller numbers are larger, such as $\frac{1}{3} > \frac{4}{7} > \frac{8}{9}$; Liu et al., 2012; Rinne et al., 2017; Stafylidou & Vosniadou, 2004). Students might also demonstrate gap thinking, erroneously believing that the smaller the 'gap' (i.e., the absolute difference between the numerator and denominator), the larger the fraction (González-Forte et al., 2020, 2023; Maricau et al., 2026). When performing operations with uncommon denominators, students may select the larger denominator for addition and the smaller denominator for subtraction (Braithwaite et al., 2017; Di Lonardo Burr et al., 2020; Xu et al., 2024). These types of errors are thought to reflect a transitory phase where students have developed some understanding that larger fraction values can consist of smaller numbers but they still erroneously view fractions as two independent natural numbers (Liu et al., 2012; Rinne et al., 2017; Stafylidou & Vosniadou, 2004).

Misconceptions can also be seen in the understanding of fraction magnitude and density. For example, students may believe that the unit is always smaller than the fraction, stating that $\frac{1}{1}$ is the smallest fraction or placing '1' at the far left when ordering fractions (Liu et al., 2012; Stafylidou & Vosniadou, 2004), or struggle to place a fraction on a 0 to 1 number line if the numerator or denominator is greater than 1 (Zhang et al., 2017). In contrast, others may believe that the unit is always bigger than the fraction, placing '1' at the far right even in the presence of improper fractions (e.g., $\frac{1}{7} < \frac{5}{6} < \frac{4}{3} < 1$; Liu et al., 2012; Stafylidou & Vosniadou, 2004). Moreover, even if students understand that a fraction is smaller than the unit when the numerator is smaller than the denominator (i.e., a proper fraction) and bigger than the unit when the numerator is larger than the denominator (i.e., an improper fraction) they may not understand the density of rational numbers. That is, students may not recognize that there is an infinite number of fractions between any two fractions (Lamon, 2020; McMullen & Van Hoof, 2020; Vamvakoussi & Vosniadou, 2010).

Overall, the multifaceted nature of fractions means that students may hold various misconceptions and make mistakes across numerous assessments. Moreover, students may make mistakes at both the beginning and intermediate stages of instruction as they continue to refine their knowledge (VanLehn, 1999). Importantly, the presence of multiple misconceptions does not imply a single coherent underlying misunderstanding; rather, errors may reflect different sources of difficulty that reflect the demands of the task. As a result, conceptual change may be characterized by instability, regression, and idiosyncratic pathways rather than smooth progression.

1.2 The Gradual Development of Fraction Knowledge and Persistence of Errors

The process of acquiring fraction knowledge is slow with instruction spread across several years. In a longitudinal study of fraction development from Grade 3 to Grade 6, Hansen et al. (2017) found that students with persistent difficulties (i.e., low-growth in fraction concepts and procedures) struggled with fraction magnitude and arithmetic, often incorrectly applying whole number knowledge. Difficulties with fractions are common and challenging to overcome, with studies showing minimal growth in some fraction concepts over several years of instruction (Grade 4 to Grade 6; Resnick et al. 2016). Moreover, Middleton et al. (2015) found that students entered middle school with several appropriate strategies for solving rational number problems but left with only slightly expanded sets of skills, often relying on inefficient ad hoc, means-end reasoning.

Consistent with these findings, cross-sectional studies highlight the persistent challenges fractions pose. For example, although whole-number bias decreased between Grade 4 and 8, it remained present for a substantial minority of students (Braithwaite & Siegler, 2018b) and has even been found in skilled adults (DeWolf & Vosniadou, 2015). Although some students appear to reorganize their initial interpretations of fractions as they progress through instruction, many continue to hold misconceptions about fraction concepts, such as viewing fractions as two independent natural numbers and only viewing fractions as part of a whole rather than as a relation between the numerator and denominator (Liu et al., 2012; Stafylidou & Vosniadou, 2004). Developmental accounts of cognition have long suggested that conceptual change may involve periods of reorganization and temporary instability rather than strictly linear improvement (Werner, 1958). In the domain of fractions, conceptual change theory suggests that learning rational numbers often requires more than enriching existing whole-number knowledge; it may involve reorganizing prior assumptions about number, magnitude, and numerical relations (Stafylidou & Vosniadou, 2004; Vosniadou, 1994). Because fractions are multifaceted, foundational concepts that appear mastered may still require extended time to fully integrate, during which new instructional content can temporarily disrupt prior understandings. From this perspective, periods of apparent regression or inconsistency in students' fraction performance may reflect active conceptual reorganization rather than simple error or lack of knowledge.

Notably, identifying where a student has gone wrong is a demanding task, as students may respond in idiosyncratic ways that do not align with a single problem-solving approach (Hansen et al., 2017; Tatsuoka, 1984). Some research has therefore moved beyond accuracy to examine underlying reasoning processes in more detail, including interview-based analyses and response pattern approaches (e.g., González-Forte et al., 2023; Halme et al., 2024; Maricau et al., 2026; Reinhold et al., 2023). These studies suggest that fraction errors are often linked to systematic reasoning patterns that may coexist with successful strategies and, in some cases, remain resistant to change or outside students' awareness. Longitudinal analyses using latent transition approaches have also begun to map developmental change in fraction understanding, showing that students may transition gradually through intermediate conceptual states and that conceptual and procedural knowledge profiles often remain relatively stable over time (Lenz et al., 2024; McMullen et al., 2015; Van Hoof et al., 2018). However, because these approaches model transitions between aggregated latent profiles, they may obscure short-term fluctuations, temporary regressions, and the coexistence of multiple strategies and misconceptions within individual learners across tasks and subconstructs. These limitations highlight the need for longitudinal approaches that track within-student changes in specific errors and reasoning patterns across multiple fraction subconstructs.

1.3 The Present Study

The present study investigated the developmental trajectories of fraction understanding by tracking the same cohort of students across four time points over two years of fraction learning (Term 1 and Term 2 of Grades 4 and 5). We examined distinct subconstructs of fractions based on the fraction topics students were expected to receive instruction on by the end of Grade 5 (Zhang & Siegler, 2022). More specifically, in the Chinese curriculum, students are introduced to foundational fraction concepts and common-denominator arithmetic in the latter half of Grade 3. By Grade 4, students are expected to have a stable grasp of these initial subconstructs, which are subsequently reinforced through decimal instruction to support the integration of rational numbers representations. This provides a platform for the transition to more advanced topics in Grade 5, such as improper fractions and operations with different denominators.

To capture possible shifts in reasoning over time, the present study examined not only overall accuracy, but also the consistency and transitions of specific response patterns across tasks and subconstructs. Specifically, the present study addressed three research questions: i) To what extent does accuracy across distinct fraction subconstructs improve over a two-year instructional period, and do these trajectories differ by subconstruct? ii) To what degree does response consistency differ across subconstructs, and does high accuracy always imply conceptual stability? iii) What specific pathways of conceptual change (i.e., transitions between errors and correct strategies) characterize the learning journey within each subconstruct?

For each subconstruct, we classified student responses according to previously documented misconceptions and errors (e.g., Braithwaite & Siegler, 2018a, 2018b; DeWolf & Vosniadou, 2015; Di Lonardo Burr et al., 2020, 2022; Mack, 1995; Ni & Zhou, 2005; Xu et al., 2024). We used a descriptive approach to investigate both the frequency of each type of response at each time point, as well as the individual patterns of change over the two-year period. Rather than treating fluctuations or shifts in responses as measurement noise, this approach considers them as potentially meaningful features of conceptual change. We anticipated that incorrect reasoning would be evident across multiple fraction assessments and that response patterns would shift over time as students progressed through instruction. By integrating descriptive trajectories with quantitative indices of stability, this work aims to provide researchers and educators with a detailed account of how fraction understanding consolidates (or fails to consolidate) across subconstructs during instruction.

2. Method

2.1 Participants

Students (T1: $N = 1{,}055$; 546 boys; $M_{\text{age}} = 9.8$ years; $SD = 0.7$) were recruited from 24 classrooms in three public elementary schools in Jinan, Shandong, China. Ethics approval was obtained from the Institutional Review Board at Shandong Normal University. After principal approval, letters were sent home by the school inviting students to participate. Parental education levels were obtained as a proxy for socioeconomic status (SES). Representative of low-middle SES in China, the median education level was a high school degree for mothers and between a high school and college degree for fathers.

2.2 Procedure

Students were tested at four time points: Grade 4 Term 1 (T1; December 2021), Grade 4 Term 2 (T2; June 2022), Grade 5 Term 1 (T3; December 2022), and Grade 5 Term 2 (T4; June 2023). Testing was completed during school hours in classrooms. Trained experimenters, who had detailed testing manuals outlining testing procedures, oversaw group testing sessions. Two experimenters were present in each classroom, with one circulating to ensure students followed the instructions and the other focusing on task administration.

2.3 Transparency and Openness

This study uses data from the Chinese Children's Mathematical Cognition, Affection and Motivation Project, a large, longitudinal project investigating the development of mathematical affect and skills. Only measures relevant to the present study (i.e., fraction assessments) are described. Data were analyzed in R Version 4.4.1 (R Core Team, 2024) using the *tidyr* (Wickham et al., 2024), *data.table* (Dowle & Scrinivasan, 2021), *ggplot2* (Wickham, 2016), *ggsankey* (Sjoberg, 2024), *emmeans* (Lenth, 2024), *lmerTest* (Kuznetsova et al., 2017), *DHARMa* (Hartig, 2024) and *lme4* (Bates et al., 2015) packages. The study's pre-registration, measures, coding scheme, and anonymized data are available at the Open Science Framework (OSF): https://osf.io/hce9v. This study's design and analysis plan were pre-registered after data had been collected but before analyses were undertaken. While the primary descriptive approach and classification of errors followed the pre-registered plan, we included additional exploratory analyses (i.e., linear and generalized linear mixed-effects models) to statistically evaluate developmental growth and response stability across subconstructs. These models, which were included to quantify longitudinal trajectories of conceptual change and response stability, are reported as exploratory and interpreted accordingly.

2.4 Measures

2.4.1 Fraction Mapping

This task was designed to assess the part-whole subconstruct. Students were presented with shapes that had been partitioned into equal-sized sections with some of the sections shaded. For each of the 14 items, students were asked to write down the fraction that corresponded to the shaded portion. Students were given one minute to complete as many items as possible, in order. Internal reliability (Cronbach's α) ranged from .92 to $>$ .99 across time points.

2.4.2 Fraction Comparison

These items assessed the measurement subconstruct. Students were presented with pairs of fractions of differing magnitude and asked to circle the larger of the two. The 24 pairs consisted of eight pairs with common numerators, eight pairs with common denominators, and eight pairs with different numerators and denominators. Of the 24 trials, 14 were congruent trials, where the larger fraction also has the larger components (e.g., $\frac{2}{5} < \frac{4}{7}; \frac{3}{5} < \frac{4}{5}$), nine were incongruent trials, where the larger fraction has the smaller components (e.g., $\frac{3}{5} > \frac{4}{9}; \frac{3}{7} > \frac{3}{9}$), and one was neither congruent nor incongruent (i.e., $\frac{4}{7} < \frac{5}{6}$). Students were given two minutes to complete as many items as possible, in order. Internal reliability (Cronbach's α) ranged from .98 to > .99 across time points.

2.4.3 General Fraction Assessment

For the general fraction assessment, students were given 10 minutes to work through 11 problems (some with multiple parts), with the option to skip and return to problems.[1] Internal reliability (Cronbach's α) ranged from .86 to .96 across time points.

Fraction Ordering. These items assessed the measurement subconstruct. Students ordered two sequences of numbers from smallest to largest: i) proper fractions (i.e., $\frac{1}{4}, \frac{1}{8}, \frac{4}{7}, \frac{8}{9}$) and ii) two proper fractions, the number 1, and an improper fraction (i.e., $\frac{2}{3}$, 1, $\frac{5}{4}, \frac{4}{9}$).

Fraction Arithmetic. These items assessed the operator and quotient subconstructs. Students solved fraction addition and subtraction problems. Four items were presented in symbolic format and four items were presented pictorially (non-symbolic; see OSF for examples). For each of the non-symbolic and symbolic sets there were two addition and two subtraction problems; half of the problems had common denominators.

---

[1] Due to an administrative error, some students received an incorrect version of one of the word problems (Problem 6) and thus it was excluded from analyses. For the number line, many students only answered part of the question (i.e., they did not provide an explanation). As such, we have moved the number line analyses to Supplementary Materials.

Abstract Understanding. Four problems asked students to think about fractions at an abstract level to assess part-whole (i.e., *Sam says that* $\frac{1}{3}$ *of the figure below is shaded. Is she correct? Why or why not?* accompanied by an unequally partitioned shape), measurement (i.e., *Round* $\frac{8}{9}+\frac{11}{12}$ *to the nearest whole number*), operator (i.e., *What is* $\frac{1}{2}$ *of* $\frac{1}{2}$*?*), and density (i.e., *How many possible fractions are between* $\frac{1}{4}$ *and* $\frac{1}{2}$*?*) concepts.

Problem Solving. These items assessed the part-whole subconstruct in a contextual setting. Students were asked to solve two story problems involving fractions: 1) *Sabrina cut a pizza into 4 equal pieces. She ate one piece. What fraction of the whole pizza did Sabrina eat?* 2) *Richard equally divided a bag of rice into 8 pieces. If he eats one piece every day, how much of the bag of rice did Richard eat after 7 days?*

2.5 Response Coding

For each measure students' raw responses were entered into a spreadsheet by the two lead authors. Scripts were then written, in *R*, to categorize each student's responses to each item based on a pre-registered coding scheme (see OSF). Categories of misconceptions were selected based on errors noted in previous research. If a response fit into more than one category, it was prioritized based on its conceptual depth (e.g., a misconception was prioritized over a simple procedural error[2]). After identifying responses that fit into the pre-defined categories, the first two authors inspected remaining uncategorized responses. If an error did not fit into a category but a strategy could be derived and occurred across at least 20 students or 2% of students (whichever was greater), additional categories were added. If an error occurred infrequently it was categorized as "mixed". For additional response coding details and inter-rater reliability, see

[2] For fraction arithmetic, if the student used the wrong operation but got the correct response, it was similarly categorized as correct.

Supplementary Materials.

When analyzing frequency of responses at each individual time point and for the mixed-effects models, all available observations were used. When examining longitudinal pathways via Sankey diagrams, only students with complete data were included to ensure the flow represented the same cohort over time.[3] In these diagrams, the width of the flows and the size of the nodes are proportional to the number of students within each response category.

3. Results

3.1 Developmental Change and Response Stability

To evaluate how fraction understanding unfolds over time, we examined changes in response categories across all subconstructs. Responses for part-whole, measurement, arithmetic/operator, and density tasks were categorized into four profiles: (i) whole-number bias (WNB); (ii) partial understanding (Part); (iii) correct (C); and (iv) mixed or blank (M/B). Detailed coding criteria for each subconstruct are provided in Table S1 in the Supplementary Materials. As shown in Figure 1, while the proportion of responses consistent with WNB declined sharply from T1 to T4, responses reflecting partial understanding of fractions remained relatively stable. Unless otherwise specified, model-based results reflect aggregate developmental trends across the sample, whereas entropy and transition analyses describe distributional stability and within-person response dynamics over time.

3.1.1 Developmental Trajectories in Accuracy

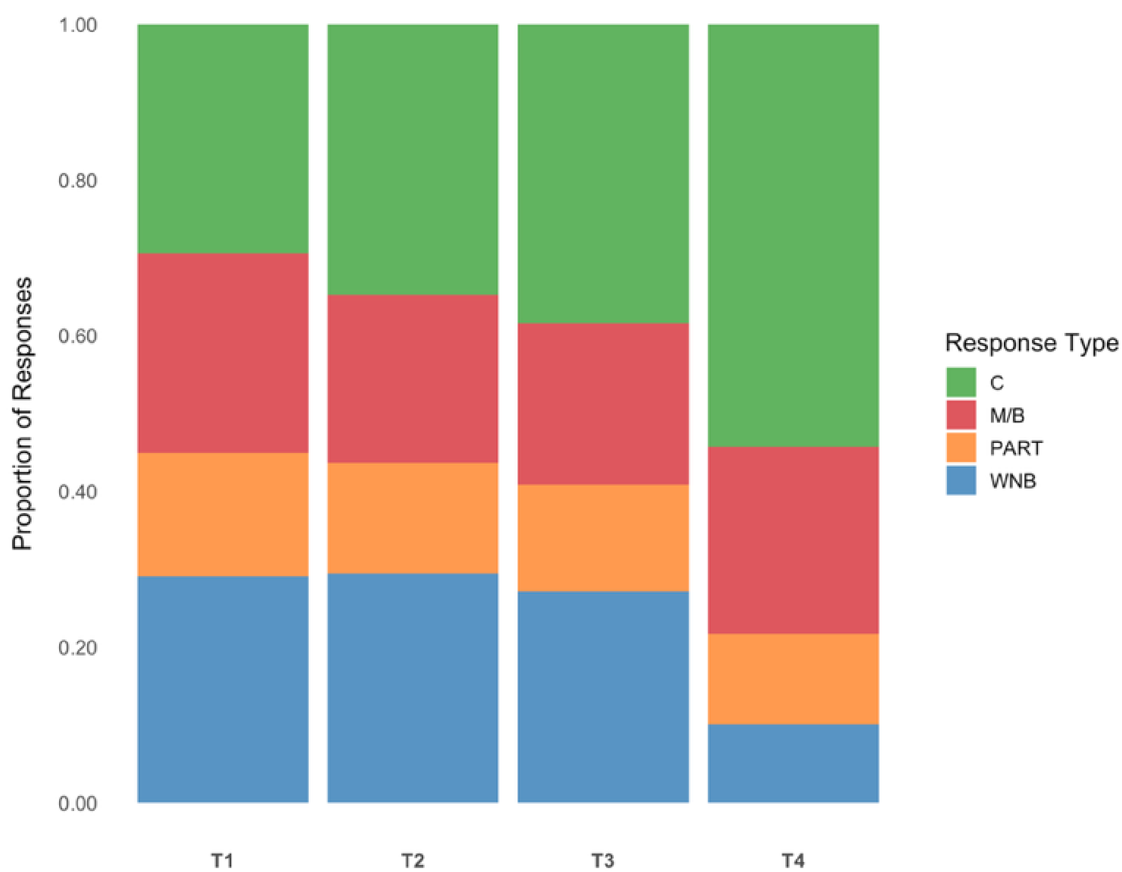
In an exploratory analysis, to examine longitudinal changes in response accuracy across fraction subconstructs, we fitted a generalized linear mixed-effects model with a binomial link

[3]In the pre-registration we stated that only students who did not provide a blank response at all four time points would be included in the transition analyses. However, for some tasks, excluding blank responses meant that approximately two-thirds of the sample were excluded from subsequent analyses. Thus, we decided to include blank responses in the Sankey diagrams.

function. The model included fixed effects of time point (categorical), subconstruct, and their interaction, with participant included as a random intercept to account for repeated observations nested within students. Degrees of freedom for mixed-effects models were estimated using Satterthwaite's approximation as implemented in the *lmerTest* package (Kuznetsova et al., 2017), and fixed effects were evaluated using Wald-type F-tests appropriate for large-sample mixed models. Model convergence was successful with no convergence warnings detected.

Figure 1

Proportion of Responses by Response Category At Each Time Point



Note. WNB = Responses consistent with whole number bias/whole number strategies or procedures; PART = Responses consistent with partial understanding; M/B = Mixed or blank responses; C = Correct responses

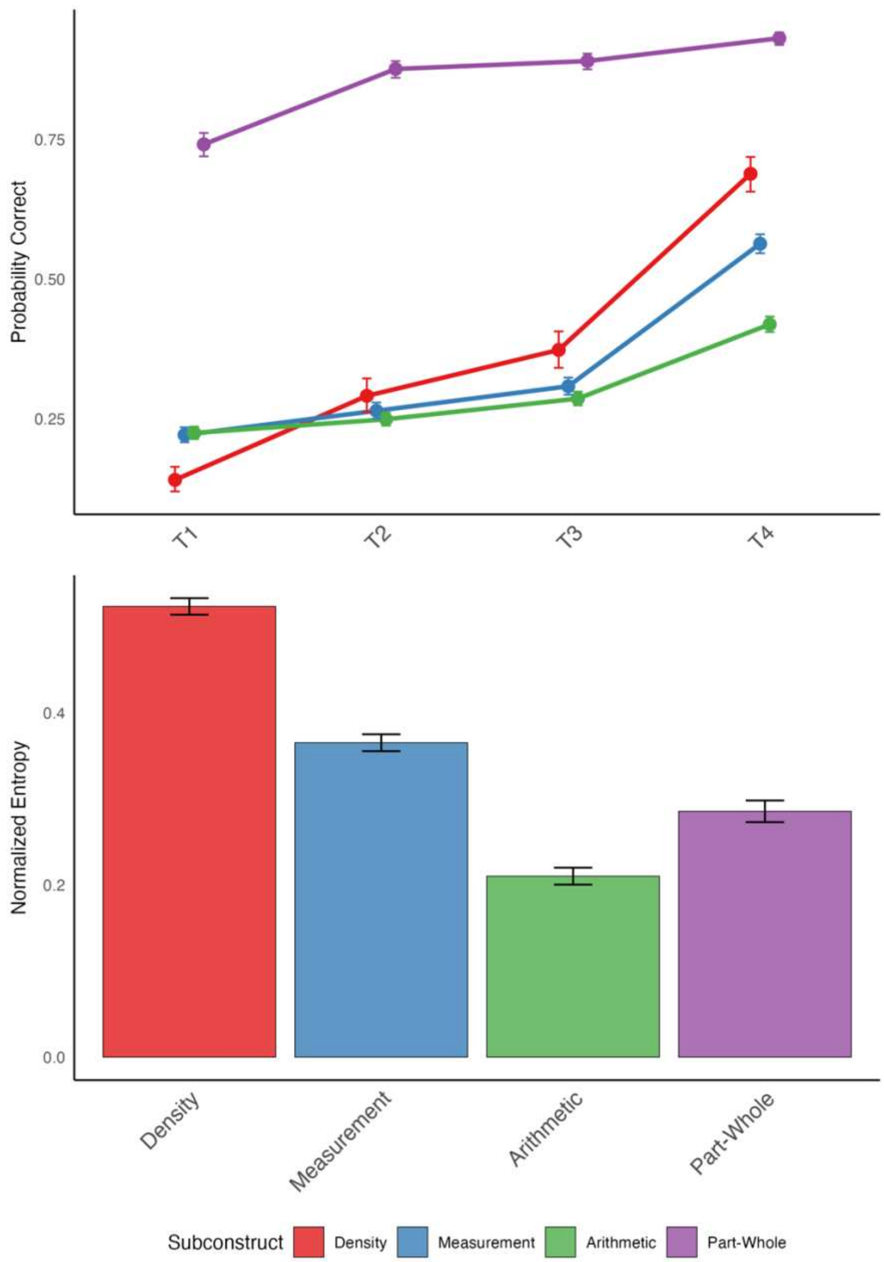
The model revealed significant main effects of time point ($F(3, 1792.9) = 597.63$, $p < .001$) and subconstruct ($F(3, 5971.7) = 1990.56$, $p < .001$), and a significant time point by subconstruct interaction, $F(9, 366.7) = 40.74$, $p < .001$, indicating that improvements in accuracy unfolded differently across subconstructs. As illustrated in Figure 2 (top panel), part-whole understanding was high at T1 (74%) and approached ceiling by T2 (88%). In contrast, density performance was poor until T4, where accuracy increased sharply from 38% to 68%. Measurement tasks showed a late-wave jump, increasing from 31% at T3 to 56% at T4. Arithmetic/operator accuracy increased more slowly, rising from 22% at T1 to 42% at T4.

3.1.2 Response Stability and Entropy

To determine whether gains in accuracy reflected stabilized understanding or shifting response patterns, we conducted an exploratory analysis examining longitudinal response inconsistency using normalized Shannon entropy. Entropy quantified variability in response categories across time points, with higher values indicating greater instability over time. Entropy and switching indices capture within-person variability in response categories across time points. Entropy values were modelled as a function of subconstruct, with mean task accuracy included as a covariate, using a linear mixed-effects model with participant included as a random intercept. Model diagnostics using the *DHARMa* package (Hartig, 2024) indicated an adequate model fit with no evidence of assumption violations. A simulation-based Kolmogorov-Smirnov test confirmed that the scaled quantile residuals did not deviate significantly from a uniform distribution ($D = 0.002$, $p = .839$) and the estimated dispersion parameter indicated no evidence of overdispersion ($\phi = 1.003$, $p = 0.574$).

Figure 2

Developmental Trajectories and Response Inconsistency Across Fraction Subconstructs



Note. Top panel displays model-estimated predicted probabilities of correct responses derived from a binomial generalized linear mixed-effects model. Bottom panel displays normalized Shannon entropy values (0-1), representing longitudinal inconsistency in response categories across time, estimated using a linear mixed-effects model with subconstruct and mean task accuracy as predictors. Error bars in both panels represent 95% confidence intervals. Lines in top panel have been jittered to reduce overlap.

The model revealed that subconstruct was a significant predictor of entropy, $F(3, 3184.8) = 815.65$, $p < .001$, while accounting for mean task accuracy, $F(1, 3107.7) = 924.90$, $p < .001$. As shown in Figure 2 (bottom panel), entropy was the highest for density ($M_{adj} = 0.52$), followed by measurement ($M_{adj} = 0.37$), and lower values for part-whole ($M_{adj} = 0.29$) and arithmetic/operator ($M_{adj} = 0.21$) tasks (see Table S2). Pairwise comparisons confirmed that all subconstructs differed significantly from one another (all ps < .001; see Table S3). Notably, arithmetic and part-whole differed in accuracy but showed comparatively smaller differences in entropy, indicating that differences in performance were not always mirrored by differences in response stability..

3.2 Qualitative Pathways of Conceptual Change

While the previous models establish the broad parameters of growth and stability, they do not capture the content of students' incorrect reasoning. To address this, following our pre-registered analytic plan, we used a response-centred descriptive approach to identify the specific errors and misconceptions defining these trajectories. Across subconstructs, a primary distinction emerged between tasks assessing part-whole understanding and those involving fraction magnitude and arithmetic. Part-whole tasks showed high initial accuracy and steady declines in error rates, reflecting relatively consistent performance. In contrast, magnitude- and arithmetic-related tasks exhibited more varied error patterns and frequent shifts between incorrect strategies. To document these patterns while minimizing redundancy, we present consolidated summary tables reporting the most frequent and second most frequent response categories at each time point for each subconstruct. These tables are paired with multi-panel Sankey diagrams illustrating individual-level transitions. Full frequency tables, rare error categories, and item-level breakdowns are reported in the Supplementary Materials.

3.2.1 Part-Whole Understanding

Students' understanding of the fundamental part-whole subconstruct was assessed through the fraction mapping, problem solving, and unequal partitions tasks, summarized in Table 1 by the most frequent and second most frequent response types at each time point. Across tasks, performance was characterized by high accuracy, early stability, and low rates of persistent misconceptions. On the fraction mapping task, students showed high consistency in strategy use, with correct responses modal at all four time points (Table 1). Misconceptions were infrequent, with whole-number bias notably absent. The most frequent error at T1 was inverting the fraction, a transient notation error that largely disappeared by T2.

Similarly, pizza and rice problems showed high accuracy at all time points (Table 1), with errors primarily reflecting misinterpretation of the question rather than conceptual misunderstanding. Individual trajectories (Figure 3) demonstrate this stability; once students reached correct responses, reversion to incorrect strategies was rare. By the end of Grade 5, nearly 90% of students accurately mapped story problems to symbolic notation.

The unequal partitions task assessed recognition that fractions must represent equal parts. Fewer than 5% of students provided an incorrect response at any time point, and approximately two-thirds responded correctly across all four waves. The most common correct rationale referenced unequal division, with transitions in Figure 3 showing movement toward this explanation over time. By T4, the majority of students could adequately justify their reasoning.

Table 1

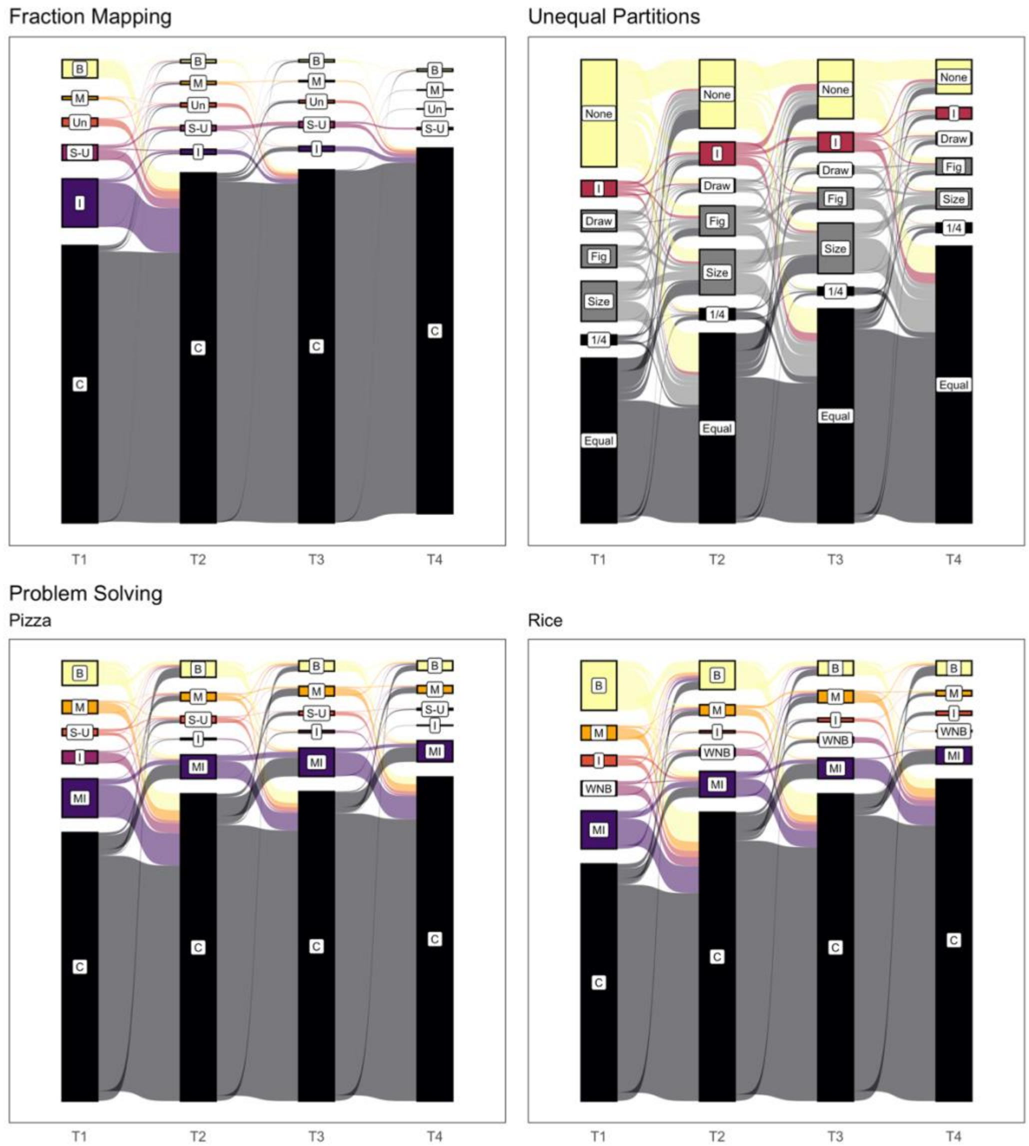
Most Frequent and Second Most Frequent Response Types Across Time Points for Part-Whole Tasks

Task	Response (%)			
	Time 1	**Time 2**	**Time 3**	**Time 4**
Fraction Mapping	**Correct (69.6)**	**Correct (89.2)**	**Correct (92.5)**	**Correct (95.7)**
	Inverting (11.6)	Blank (3.0)	Blank (1.9)	Mixed (1.9)
Problem Solving (Pizza)	**Correct (73.9)**	**Correct (84.4)**	**Correct (85.1)**	**Correct (89.2)**
	Misinterpreted (10.3)	Misinterpreted (6.6)	Misinterpreted (7.8)	Misinterpreted (5.7)
Problem Solving (Rice)	**Correct (65.6)**	**Correct (79.6)**	**Correct (84.5)**	**Correct (88.4)**
	Blank (13.4)	Misinterpreted (7.0)	Misinterpreted (5.4)	Misinterpreted (4.8)
Unequal Partitions (Response)	**Correct (78.5)**	**Correct (85.7)**	**Correct (87.3)**	**Correct (93.9)**
	--	--	--	--
Unequal Partitions (Rationale)	**Unequal Division (44.0)**	**Unequal Division (50.6)**	**Unequal Division (56.6)**	**Unequal Division (72.4)**
	No Rationale (28.3)	No Rationale (17.6)	No Rationale (15.2)	No Rationale (9.1)

Note. For the unequal partitions response students could only respond with correct, incorrect, or blank. Most frequent response is bolded. Second most frequent response is in regular text. Complete response distributions and all coded error categories for these tasks can be found in Tables S4 and S5.

Figure 3

Proportion and Transition in Category of Response for Part-Whole Assessment



Note. Top left panel shows transition patterns for fraction mapping ($N = 954$). Bottom panels show transition patterns for problem solving ($N = 950$). C = Correct; MI = Misinterpreted; WNB = Whole number bias; I = Inverted; S-U = Added shaded for numerator and unshaded for denominator; Un = Unshaded; M = Mixed; B = Blank.

Top right panel shows transition patterns for unequal partitions explanations ($N = 950$). C = Correct; I = Incorrect; B = Blank; Equal = Unequal division; 1/4 = One-fourth; Size = Size of partitions; Fig = Shapes, parts, and lines of figure; Draw = Drawing of correct partition; None = No explanation provided.

3.2.2 Measurement and Magnitude Understanding

Students' understanding of fraction magnitude (i.e., measurement interpretation) was assessed through comparison, ordering, and estimation tasks. In contrast to the early stability seen in part-whole tasks, these assessments revealed a transitional profile marked by persistent whole-number bias and frequent shifts between incorrect strategies. As shown in Table 2, although approximately 50% of students could accurately compare fractions by T3, fewer than 10% could correctly order fractions. The increase-increase error was the most prevalent misconception, reflecting independent consideration of numerators and denominators. Figure 4 illustrates high volatility in these tasks, with frequent switching between whole-number-based strategies and reversion from correct to incorrect responses. By T4, a growing mixed category (22.7% for ordering) reflected attempts to apply more complex but incomplete procedures, such as finding common denominators or converting to decimals.

Additional evidence of persistent whole-number bias was observed in the task requiring students to round $\frac{8}{9}+\frac{11}{12}$ to the nearest whole number (Table 2). Fewer than 10% of students responded correctly at the first three time points, with most errors involving treating the numerators and denominators as independent whole numbers or performing literal operations on them (e.g., $\frac{19}{21}$). Figure 4 illustrates that the most frequent transition patterns involved remaining in or switching between these two error categories. Over half of students (53%) failed to provide a correct response even once across the two-year period, with meaningful gains emerging only at T4.

Table 2

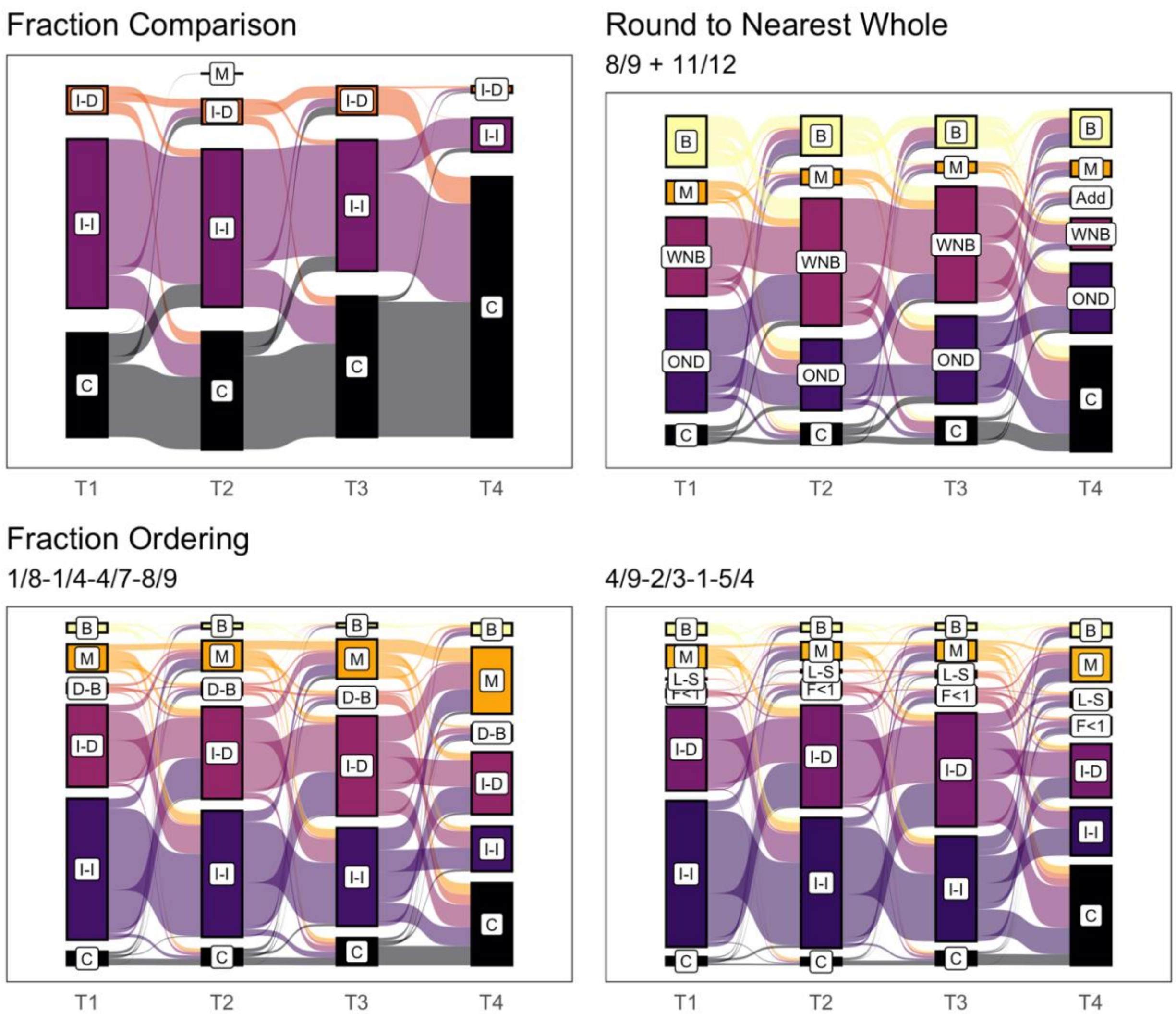
Most Frequent and Second Most Frequent Response Types Across Time Points for Measurement Tasks

Task	**Response (%)**			
	Time 1	**Time 2**	**Time 3**	**Time 4**
Comparison	**Increase-Increase (53.1)**	**Increase-Increase (47.8)**	**Correct (50.8)**	**Correct (88.9)**
	Correct (38.0)	Correct (43.1)	Increase-Increase (39.2)	Increase-Increase (8.7)
Ordering $\frac{1}{4}, \frac{1}{8}, \frac{4}{7}, \frac{8}{9}$	**Increase-Increase (49.2)**	**Increase-Increase (44.0)**	**Increase-Decrease (35.0)**	**Correct (29.6)**
	Increase-Decrease (29.6)	Increase-Decrease (32.5)	Increase-Increase (34.9)	Increase-Decrease (21.9)
Ordering $\frac{2}{3}, 1, \frac{5}{4}, \frac{4}{9}$	**Increase-Increase (51.1)**	**Increase-Increase (45.7)**	**Increase-Decrease (39.7)**	**Correct (34.9)**
	Increase-Decrease (30.6)	Increase-Decrease (36.6)	Increase-Increase (37.5)	Increase-Decrease (19.2)
Round Nearest Whole	**OND (37.9)**	**WNB (46.4)**	**WNB (41.9)**	**Correct (38.0)**
	WNB (29.1)	OND (26.6)	OND (32.1)	OND (25.7)

Note. Most frequent response is bolded. Second most frequent response is in regular text. WNB = Whole number bias; OND = Operations on numerator and denominator. See Supplementary Materials for details on fraction comparison categorization. Complete response distributions and all coded error categories for these tasks can be found in Tables S6 and S7.

Figure 4

Proportion and Transition in Category of Response for Measurement Assessments



Note. Top left panel shows transition patterns for fraction comparison (N = 933). Bottom panels show transition patterns for fraction ordering (N = 950). C = Correct; I-I = Increase-increase; I-D = Increase-decrease; D-B = Distance between numerator and denominator; F < 1 = Fractions are less than whole numbers; L-S = Largest-to-smallest; M = Mixed; B = Blank.

Top right panel shows transition patterns for rounding to nearest whole number (N = 950). C = Correct; OND = Operations on numerator and denominator; WNB = Whole number bias; Add = Added instead of rounding; M = Mixed; B = Blank.

3.2.3 Fraction Arithmetic and Operators

Arithmetic knowledge was assessed using common denominator, different denominator, and operator ($\frac{1}{2}$ of a $\frac{1}{2}$) tasks. Unlike measurement tasks, arithmetic errors were highly stable, with students consistently applying the same incorrect procedures across time points. By T4, most students could accurately add and subtract fractions with common denominators (Table 3). At T1, however, 40-50% performed operations on both the numerator and denominator, an error consistent with whole number bias. Interestingly, while this was the only classifiable error for symbolic items, non-symbolic items elicited part-whole errors (e.g., inverting answers; miscounting shaded/unshaded pieces), suggesting students approached symbolic and visual representations using different strategies. Figure 5 illustrates that once mastery was achieved, it was highly stable; nearly all students (95-99%) provided a correct response at least once, and approximately one-third were correct at all four time points.

Different denominator arithmetic proved most difficult, with over 99% of students making errors at the first three time points (Table 3). Meaningful gains emerged only at T4, alongside a new error involving incorrect use of common denominators. Students were likely attempting to apply new procedural instruction without full conceptual understanding. For symbolic addition, 38% of students made the same error at every time point, with nearly half (48%) consistently applying operations on the numerator and denominator. For subtraction and non-symbolic items, students frequently cycled between operations on the numerator and denominator and simply selecting the bigger denominator. The qualitative stability shown in Figure 6 is consistent with the entropy model, which identified arithmetic as a domain of high response consistency despite low overall accuracy.

Table 3

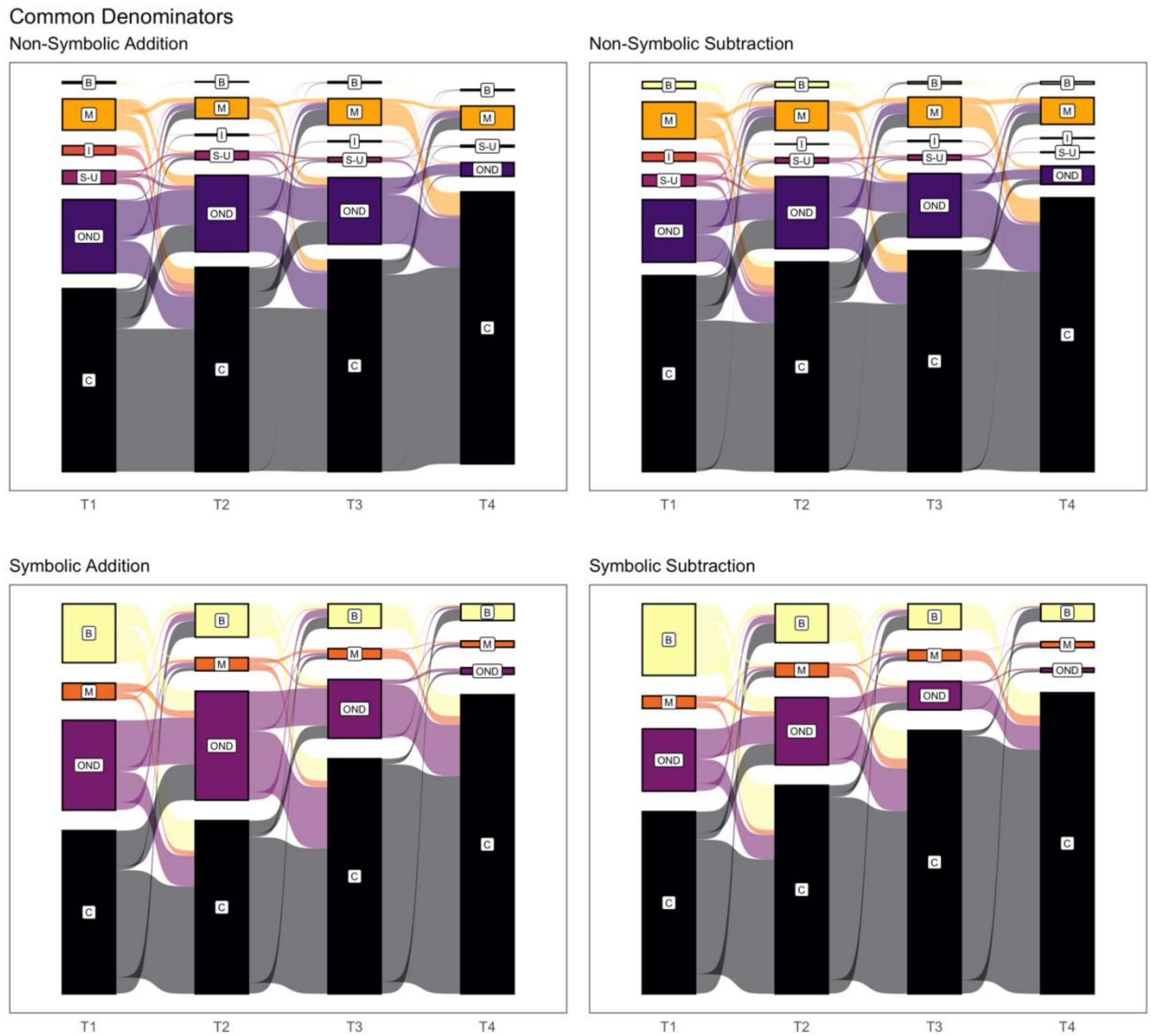
Most Frequent and Second Most Frequent Response Types Across Time Points for Arithmetic Tasks

Task	Symbolic Arithmetic Response (%)				Non-Symbolic Arithmetic Response (%)			
	Time 1	**Time 2**	**Time 3**	**Time 4**	**Time 1**	**Time 2**	**Time 3**	**Time 4**
Common Denominator Addition	**Correct (48.6)**	**Correct (53.1)**	**Correct (72.1)**	**Correct (90.8)**	**Correct (58.3)**	**Correct (66.1)**	**Correct (68.5)**	**Correct (86.9)**
	OND (28.0)	OND (32.9)	OND (17.4)	Blank (5.2)	OND (23.7)	OND (24.3)	OND (20.9)	Mixed (7.7)
Common Denominator Subtraction	**Correct (54.6)**	**Correct (63.9)**	**Correct (80.4)**	**Correct (91.7)**	**Correct (60.6)**	**Correct (65.9)**	**Correct (68.5)**	**Correct (84.4)**
	Blank (22.0)	OND (20.0)	OND (8.7)	Blank (5.4)	OND (19.2)	OND (21.4)	OND (19.6)	Mixed (8.6)
Different Denominator Addition	**OND (68.9)**	**OND (74.1)**	**OND (74.7)**	**Correct (44.2)**	**OND (46.4)**	**OND (44.7)**	**OND (52.9)**	**BD (24.4)**
	Blank (21.4)	Blank (11.5)	Blank (9.0)	OND (31.0)	BD (27.6)	BD (30.6)	BD (25.6)	Correct (22.1)
Different Denominator Subtraction	**Blank (38.9)**	**OND (38.7)**	**OND (40.7)**	**Correct (42.5)**	**BD (43.9)**	**BD (43.3)**	**BD (42.0)**	**BD (39.3)**
	OND (36.2)	Blank (21.5)	Blank (17.0)	Blank (11.9)	OND (28.6)	OND (25.4)	OND (27.7)	Correct (17.4)

Note. Most frequent response is bolded. Second most frequent response is in regular text. OND = Operations on numerator and denominator; BD = Bigger denominator. Complete response distributions and all coded error categories for these tasks can be found in Tables S8 and S9.

Figure 5

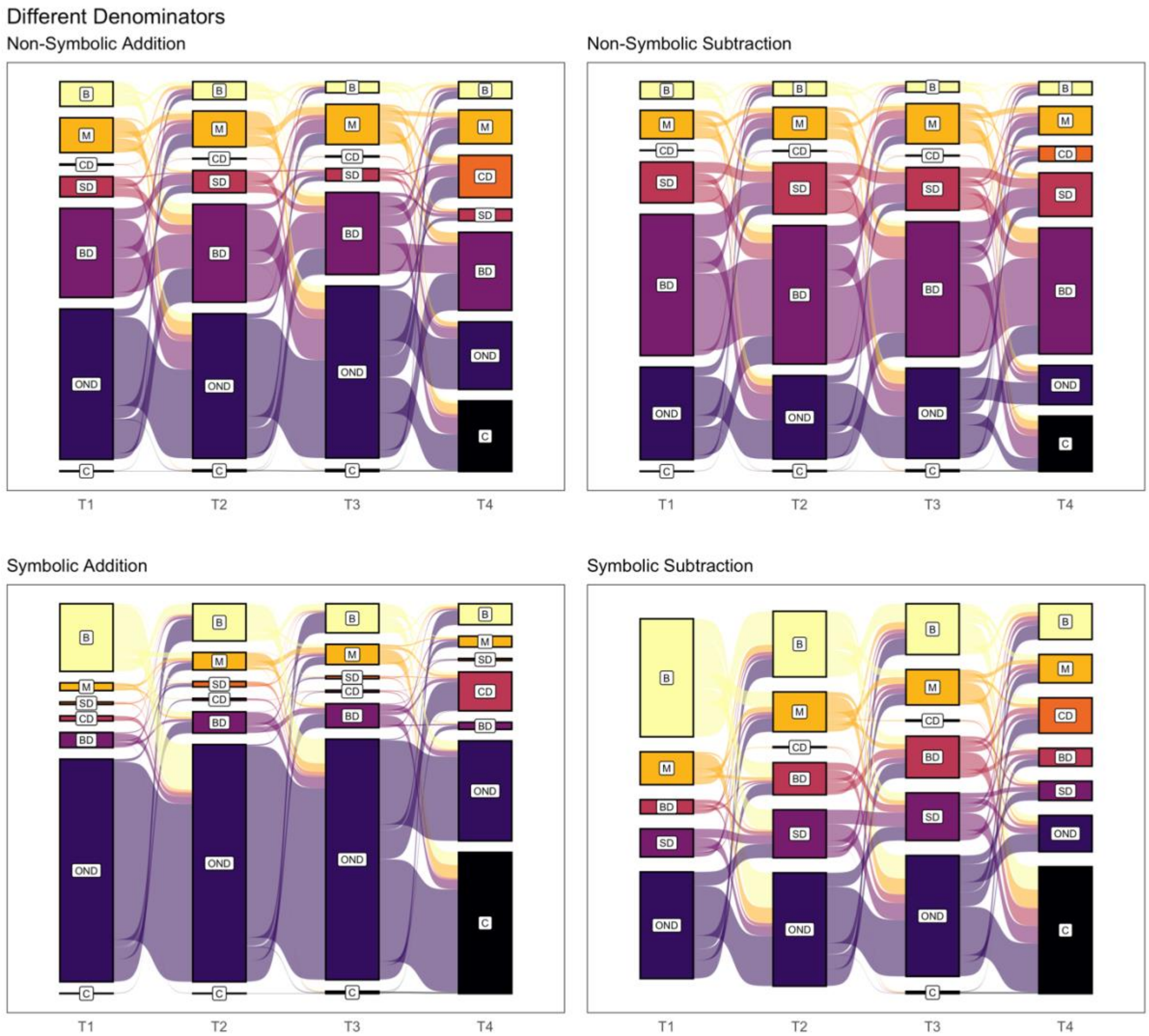
Proportion and Transition in Category of Response for Common Denominator Arithmetic



Note. N = 950. C = Correct; OND = Operations on numerator and denominator; S-U = Added shaded for numerator and Unshaded for denominator; I = Inverted; M = Mixed; B = Blank.

Figure 6

Proportion and Transition in Category of Response for Different Denominator Arithmetic



Note. N = 950. C = Correct; OND = Operations on numerator and denominator; BD = Bigger denominator selected; SD = Smaller denominator selected; CD = Found common denominator but incorrectly adjusted numerator; M = Mixed; B = Blank.

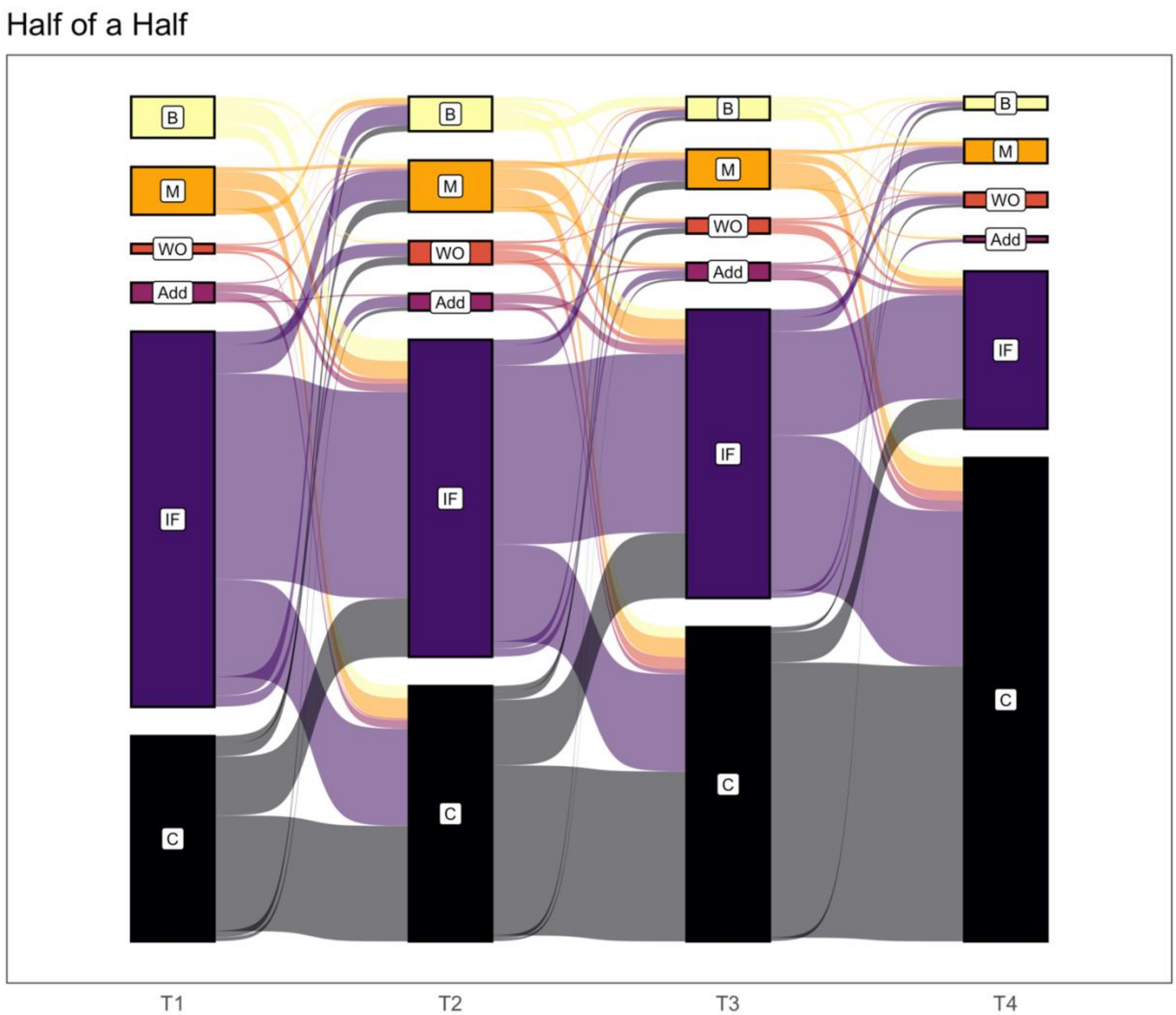
The operator task revealed a persistent independent fractions misconception, where students treated the two halves as unrelated entities. Correct responses rose from 29% at T1 to 69% at T4, while independent fractions remained the primary misconception across time points (ranging from 53% at T1 to 22% at T4; see Table S10). By the end of the study, 30% of students still struggled to conceptualize the "of" operator. Figure 7 shows that errors were highly consistent: over half of students who made an error repeated that error type over time.

3.2.4 Density Understanding

Density understanding was assessed by asking how many fractions lie between $\frac{1}{4}$ and $\frac{1}{2}$. The dominant early misconception involved treating fractions as countable, with students providing finite whole number responses. The countable fractions error accounted for 30% of responses at T1 and remained the modal error until T4. At T4, accuracy increased sharply from 38% to 68%, as students moved away from misconceptions consistent with whole number bias (see Table S11 for full distribution of errors). Figure 8 illustrates this late-breaking transition. Although most students (79%) responded correctly at least once, only 7% were correct at all time points. The dominant longitudinal pattern was a direct transition from the countable misconception to the correct response at T4.

Figure 7

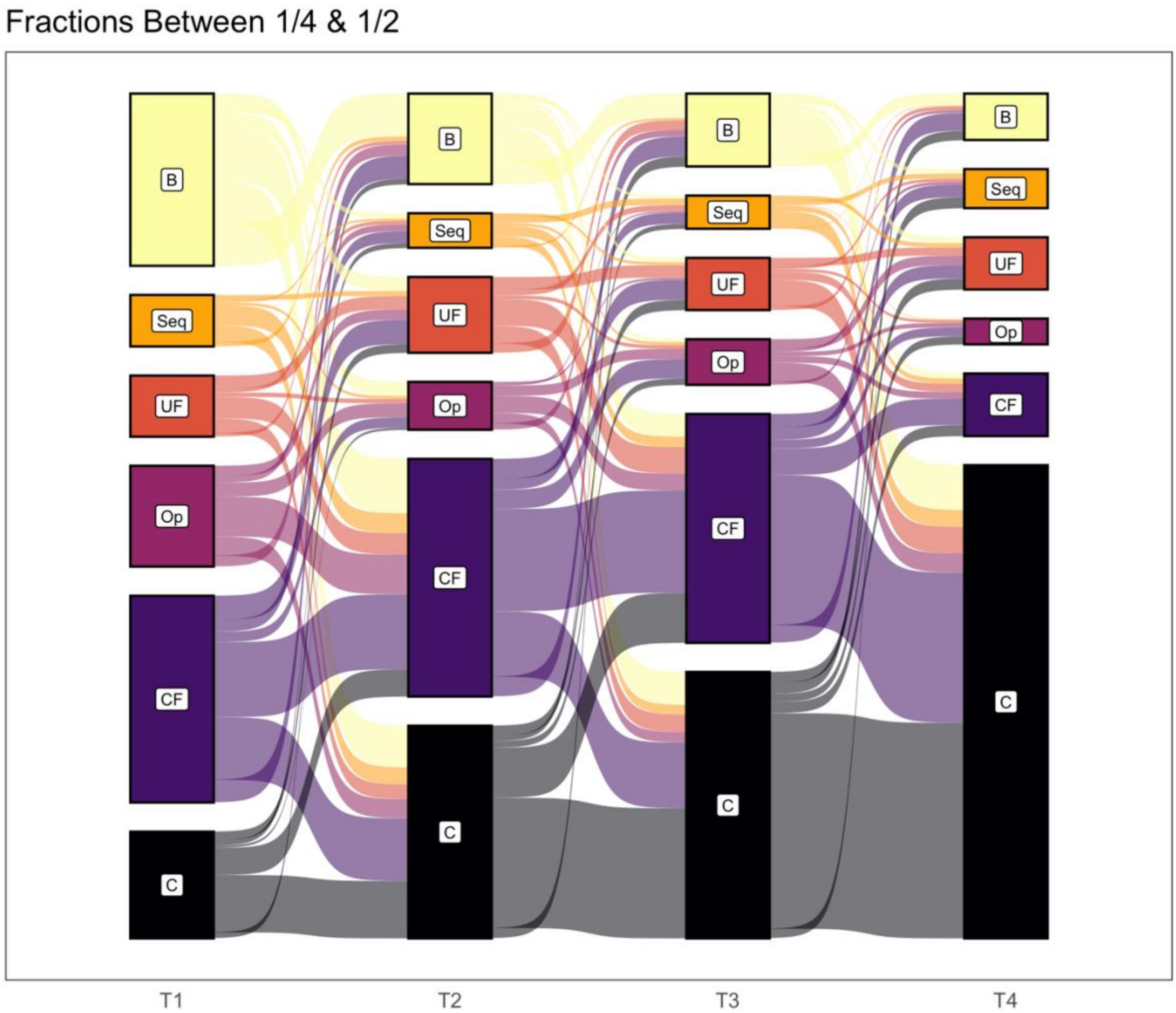
Proportion and Transition in Category of Response for Half of a Half



Note. $N = 950$. C = Correct; IF = Independent fractions; Add = Added two halves; WO = Whole only; M = Mixed; B = Blank.

Figure 8

Proportion and Transition in Category of Response for Fractions Between $\frac{1}{4}$ and $\frac{1}{2}$



Note. *N* = 950. C = Correct; CF = Countable Fractions; Op = Performed an operation; UF = Unit fractions only; Seq = Sequence of fractions; B = Blank.

3.3 Summary of Results

Across all subconstructs, students demonstrated significant growth in accuracy over time; however, this growth differed markedly in its stability. The exploratory mixed-effects models revealed substantial differences in longitudinal volatility, with some subconstructs characterized by early consolidation of understanding and others by prolonged conceptual instability. Part-whole tasks showed high initial performance and early stabilization, with most students maintaining consistent strategies over time. Arithmetic tasks, by contrast, were dominated by stable but often incorrect strategies, consistent with relatively stable error patterns over time. Measurement tasks occupied an intermediate position, exhibiting frequent shifts among multiple incorrect and partially correct strategies before convergence. Density stood apart as a threshold construct, showing minimal early progress followed by a sharp increase in accuracy accompanied by strategic reorganization at later time points. Overall, these findings indicate that similar gains in accuracy can reflect qualitatively different developmental pathways across fraction subconstructs.

4. Discussion

Developing fraction knowledge is a long, challenging process. To support students' development, we need to identify misconceptions and understand how they change over time. Research on misconceptions predominantly focuses on whole number bias (e.g., Braithwaite & Siegler, 2018b; Gabriel et al., 2023; Ni & Zhou, 2005; Van Hoof et al., 2017) or examining misconceptions within one or two fraction assessments (e.g., Di Lonardo Burr et al., 2020, 2022; Gómez & Dartnell, 2019). Far less work has examined how a broad range of fraction errors unfolds longitudinally within the same group of students. The present study investigated the progression of fraction misconceptions and errors over two years, moving beyond accuracy to examine how response patterns vary across time, tasks, and subconstructs. Overall, students showed improvement in fraction knowledge across all subconstructs, but this improvement differed in its form and stability. The observed trajectories suggest that increases in accuracy can arise from different underlying response dynamics, including stable performance, gradual change, and intermittent inconsistency.

4.1 Subconstruct Divergence: Foundational Stability versus Magnitude Volatility

The most striking contrast in our data is between the early mastery of part-whole meanings and the persistent, unstable struggle with magnitude. While many of the errors we captured have been documented in prior cross-sectional work, our longitudinal design reveals that these misconceptions do not follow an orderly, stage-like progression. Instead, the observed variation reflects task-dependent shifts within the same individuals. The fact that students employ disparate strategies and produce different errors across subconstructs suggests that early fraction understanding is highly context-dependent rather than a unified conceptual framework.

We first document a stable milestone: the mastery of part-whole interpretations. For simpler tasks that tap into part-whole meaning – a construct emphasized during early fraction instruction that builds on students' existing understanding of natural numbers (Ni, 1999) – we found that few students held misconceptions. Consistent with previous studies with students of the same age (Di Lonardo Burr et al., 2022; Xu et al., 2024), most students were accurate in fraction-to-symbol mapping, including in story problems. When misconceptions did occur, the most common errors, such as inverting and part-whole misinterpretation, suggested that some students understood that fractions consist of two parts, but they did not fully grasp the roles of the numerator and denominator. Whole number bias was rare in these tasks, appearing only in one story problem, and decimal responses were absent. Moreover, beyond just mapping fractions to symbols, most students (ranging from 79% at T1 to 94% at T4) understood that the "parts" of a fraction need to be equal. When presented with an unequally partitioned representation, students' explanations, which referred to the need for equal division, asymmetry, parts or lines missing, and so on, were consistent with part-whole interpretation. Overall, after two years of fraction instruction, students showed mastery of the part-whole meaning of fractions, representing a stable, early-acquired foundation that sharply contrasts with the later-emerging subconstructs.

Tasks tapping into measurement and magnitude showed clear evidence of natural number knowledge persistently interfering with fraction understanding. Consistent with prior research, the most common errors reflected whole number bias, with students selecting fractions with larger components for comparison, (Bonato et al., 2007; DeWolf & Vosniadou, 2015; Meert et al., 2010; Di Lonardo Burr et al., 2022) and ordering fractions by ascending numerator or denominator values (Malone & Fuchs, 2017). For arithmetic tasks, students used whole number procedures, such as adding/subtracting both the numerators and denominators, or they chose the larger denominator in arithmetic (Braithwaite et al., 2017, 2018; Di Lonardo Burr et al., 2020). Although the frequency of these errors declined, they remained evident at all time points.

However, whole number bias did not account for all errors. Many students provided responses that suggested a partial understanding of the measurement meaning of fractions. For example, many students believed that a fraction's magnitude increases as its numerator or denominator decreases (Fazio et al., 2016; Liu et al., 2012; Stafylidou & Vosniadou, 2004). Other misconceptions in fraction ordering reflected gap thinking (González-Forte et al., 2023; Pearn & Stephens, 2004) wherein students compare the difference between the numerator and denominator and then order the fractions according to those distances. With fraction arithmetic with different denominators, some students selected the smaller denominator in their response (Xu et al., 2024). Across these types of error, students are continuing to view fractions as two independent numbers, but in ways that were not directly traceable to whole number magnitude, instead reflecting transitional reasoning described in prior work (Fazio et al., 2016; Liu et al., 2012; Stafylidou & Vosniadou, 2004).

For some fraction tasks it is possible for students to derive correct solutions using whole number procedures without accessing fraction magnitude (e.g., cross-multiplication in comparison tasks; common-denominator procedures in arithmetic). Other tasks, such as fraction ordering and rounding, require direct access to magnitude. It is in these tasks that instability persisted most clearly, even when overall accuracy improved. For example, when representing half of a half, more than 40% of students across the first three time points treated the two halves as independent, drawing diagrams with two of four parts shaded. Although many students recognized that the whole should consist of four parts, they struggled to identify the correct part, effectively producing responses equivalent to $\frac{1}{2} \times \frac{1}{2} = \frac{2}{4} = \frac{1}{2}$. Poor performance on these tasks aligns with prior findings (Smith et al., 2005) and highlights the difficulty of reorganizing magnitude-based reasoning. Notably, later instruction sometimes increased instability, as students attempted to apply newly learned procedures (e.g., finding common denominators) to conceptual problems where they were not appropriate.

Understanding fraction magnitude also requires coordinating fractions with whole numbers. In the fraction ordering task, students frequently placed ‘1’ at the end of the sequence, treating it as the largest value. Across time points, this occurred in 55% of responses, consistent with frameworks in which fractions are strictly viewed as parts of a unit (Stafylidou & Vosniadou, 2004). This misconception may stem from early instructional emphases on part-whole interpretations. To develop fractions as measure, students must learn to select and construct appropriate units (Harel & Confrey, 1994; Vamvakoussi, 2015).

Comprehending the density property of fractions was also challenging. Consistent with prior research showing poor density knowledge into adolescence (McMullen & Van Hoof, 2020; Vamvakoussi et al., 2011), fewer than 40% of students at the first three time points stated that there were an infinite number of fractions between $\frac{1}{4}$ and $\frac{1}{2}$. Many attempted to count fractions or assumed that only unit fractions existed between benchmarks (e.g., $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$ ...). Density was also the subconstruct with the greatest late-stage reorganization, suggesting that conceptual change in this domain may be delayed but substantial once it occurs.

In summary, over two years of instruction, students simultaneously held multiple misconceptions and errors, with few errors appearing consistently across tasks. Fraction arithmetic responses revealed that even within a single task at a given time point, students employed different strategies depending on representation or operation. While some errors reflected whole number bias, many others reflected partial, unstable understandings. Students mastered part-whole interpretations early, but magnitude-related reasoning remained fragile and highly variable.

4.2 Fraction Misconceptions and Errors Over Time

The longitudinal design allowed us to examine how errors change across instruction. Although performance improved for all tasks, rates of improvement varied, and we observed hundreds of distinct response trajectories, highlighting extreme heterogeneity. Prior models propose orderly progressions from whole number interpretations to relational fraction understanding (e.g., Liu et al., 2012; Stafylidou & Vosniadou, 2004). Our findings complicate these accounts by showing that students often move back and forth between correct and incorrect responses rather than progressing in a linear fashion. This pattern aligns with classic developmental frameworks (e.g., Werner, 1948) and conceptual change theory of fraction learning (Stafylidou & Vosniadou, 2004; Vosniadou, 1994), which view temporary regression not as instructional failure but as part of cognitive reorganization as learners work to differentiate and integrate complex, multifaceted concepts. Thus, capturing correctness at a single time point does not necessarily reflect stable understanding; rather, it may represent a transient state on an unstable learning path.

This inherent instability cautions against grouping students into performance-based profiles. Although prior work has statistically classified students into achievement groups to capture individual differences (Hallett et al., 2010; Karamarkovich & Rutherford, 2019; Resnick et al., 2016), such groupings can obscure the heterogeneity in underlying misconceptions. For example, Karamarkovich and Rutherford (2019) identified a class of students characterized by their tendency to make few errors, but the types of errors these students made had little in common. Similarly, our detailed exploration of response patterns showed that students with similar accuracy levels often differed substantially in the types of errors they produced. Our findings suggest that accuracy profiles can be useful for broad categorization, but they risk masking meaningful differences in a students' actual conceptual state.

One of the challenges of capturing misconceptions at the individual level is that students sometimes provide responses that are unclassifiable. For difficult tasks, such as fraction ordering and different-denominator arithmetic, these unclassifiable responses increased over time. Siegler's (1996) overlapping waves theory provides a useful framework here, positing that children employ a variety of competing strategies simultaneously, both during brief transitions and over prolonged periods of development (Chen & Siegler, 2000). In the present study, the growth in unclassifiable responses reflects this competition. Closer examination showed that students were struggling to integrate new instructional information, particularly around common denominators, decimals, and units of measurement, into their existing fraction representations. Consistent with findings in algebraic problem solving (Booth et al., 2014), we highlight that specific errors become more prominent as students navigate different subconstructs at various stages of instruction. Thus, it is critical to identify periods in the curriculum where newly acquired mathematical knowledge may temporarily destabilize existing strategies before a more advanced understanding takes hold.

4.3 Educational Implications

A primary contribution of our longitudinal mapping is the evidence that fraction misunderstandings are both diverse and deeply held, often occurring simultaneously across interrelated subconstructs. Critically, students are not always aware that they hold misconceptions, reporting high confidence in their incorrect responses (González-Forte et al., 2023; Xu et al., 2024). Thus, it is essential for educators to not only identify errors but actively assist in their refutation. By systematically classifying errors across assessments, educators can develop targeted conceptual explanations to address specific misunderstandings. Interactive classroom sessions focused on collaborative error analysis have proven particularly beneficial in this regard (Metcalfe et al., 2025).

Furthermore, our findings suggest that student performance should be interpreted across multiple task formats rather than inferred from a single task. Strategic choices and error patterns vary depending on representational format (e.g., symbolic vs. non-symbolic) and task demands, meaning that single-item or single-format assessments may not fully reflect a student's underlying understanding. This is evident in our data, where fraction comparison performance improved by Time 4, whereas performance on fraction ordering remained comparatively low. One possible explanation is that students may rely on procedures such as cross-multiplication that support success in specific comparison tasks but do not transfer reliably to other formats. As a result, reliance on a single task type risks overlooking these task-specific strategies and may overstate the extent of students' conceptual understanding of fraction magnitude. More broadly, this highlights the importance for both researchers and practitioners of avoiding inferences about students' overall fraction understanding from performance on a single task. When designing targeted learning interventions or developing tutoring protocols, it may therefore be more informative to consider patterns of responses across multiple items and task formats, rather than attempting to assign students to stable ability profiles based on isolated indicators of performance.

The observed variability in student transitions suggests that traditional, one-size-fits-all classroom feedback is limited in addressing the range of misconceptions observed over time. Although learning from errors can facilitate knowledge retention and self-regulated learning (Zhang & Fiorella, 2023), effective learning still depends on timely and specific feedback to bridge the gap (Alfieri et al., 2011). However, the diversity and frequency of error types in our data highlight the potential challenges of providing consistent individualized feedback in typical classroom settings. This aligns with intervention studies in fraction learning showing that while targeted instruction and adaptive feedback can improve overall performance, effects on specific misconception reduction are often mixed (Demedts et al., 2024, 2025; D'Erchie et al., 2025).

Within this context, our findings suggest a more specific implication for instructional design: the effectiveness of feedback may depend not only on whether an error is detected, but also on whether it reflects a stable misconception or a transient strategy that varies across tasks and time. This distinction helps explain why some interventions improve accuracy without consistently eliminating underlying error patterns. It also complements prior work on computer-based tutoring systems for fraction learning (Di Lonardo Burr et al., 2020; Spitzer & Moeller, 2022) by highlighting the need to consider stability of reasoning when interpreting intervention effects. Overall, the present study suggests that fraction instruction and feedback design should account not only for the presence of misconceptions, but also for their longitudinal instability within learners.

4.4 Limitations and Future Research

In this study, we identified errors across numerous assessments based on raw student responses. By relying on product data some responses could not be classified. Although the single responses in the unclassifiable categories occurred less than 2% of the time, capturing these remains important for future interventions. Thus, in future studies, we recommend incorporating strategy reports to better categorize idiosyncratic reasoning.

Our students were enrolled in schools in China. Because instruction and curriculum heavily influence mathematical development, it is possible that our findings will not generalize to all student populations. However, many of the misconceptions were consistent with our pre-registered coding scheme, which drew from global research. Nevertheless, it is important for researchers to conduct replication studies with students educated in other countries to adequately capture misconceptions. Relatedly, we did not directly measure classroom instruction or curriculum sequencing at the individual or classroom level. As a result, we cannot directly model the influence of instructional changes (e.g., introduction of decimal topics) on observed longitudinal shifts in misconceptions. The findings therefore reflect developmental changes in students' fraction understanding within a naturalistic curriculum context, rather than being attributable to specific instructional interventions.

Finally, the use of group testing and strict time limits may have restricted students' ability to explore diverse solution strategies. It should be noted, however, that timed, high-stakes testing is a standard and ecologically valid component of the Chinese educational experience. While this format provides a realistic window into how students perform under typical classroom pressures, it created a challenge for the present study in distinguishing between blank responses caused by time constraints versus those stemming from conceptual uncertainty. More generally, because each task targets a specific aspect of fraction reasoning, performance should be interpreted as task-specific conceptual performance under timed conditions rather than as a pure measure of underlying knowledge independent of context. Future research might employ untimed assessments or a greater number of items per subconstruct to further minimize these confounding factors.

4.5 Conclusion

We examined student errors across various fraction subconstructs over two years, revealing that students simultaneously hold independent misconceptions across interrelated areas. Using an exploratory modelling approach to evaluate response entropy, we found a clear difference between stable error patterns in arithmetic and volatile, high-entropy transitions in magnitude tasks. These results show that correctness can be a transient state rather than a sign of stable understanding, as students frequently revert to earlier errors when encountering new rational number topics. Consequently, tracking these longitudinal response patterns is essential for understanding how students navigate the curriculum. This methodical classification of errors provides a framework for future research and the development of diagnostic tools that target the root of these shifting misconceptions.

References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*(1), 1–18. https://doi.org/10.1037/a0021017

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Behr, M. J., Harel, G., Post, T. R., & Lesh, R. (1993). Rational numbers: Toward a semantic analysis-emphasis on the operator construct. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds), *Rational numbers: An integration of research* (pp. 13–47). Lawrence Erlbaum Associates.

Bonato, M., Fabbri, S., Umiltà, C., & Zorzi, M. (2007). The mental representation of numerical fractions: Real or integer? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1410–1419. https://doi.org/10.1037/0096-1523.33.6.1410

Booth, J. L., Barbieri, C., Eyer, F., & Pare-Blagoev, E. J. (2014). Persistent and pernicious errors in algebraic problem solving. *The Journal of Problem Solving*, *7*(1). https://doi.org/10.7771/1932-6246.1161

Braithwaite, D. W., Pyke, A. A., & Siegler, R. S. (2017). A computational model of fraction arithmetic. *Psychological Review*, *124*(5), 603–625. https://doi.org/10.1037/rev0000072

Braithwaite, D. W., & Siegler, R. S. (2018a). Children learn spurious associations in their math textbooks: Examples from fraction arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(11), 1765–1777. https://doi.org/10.1037/xlm0000546

Braithwaite, D. W., & Siegler, R. S. (2018b). Developmental changes in the whole number bias. *Developmental Science*, *21*(2), e12541. https://doi.org/10.1111/desc.12541

Braithwaite, D. W., Tian, J., & Siegler, R. S. (2018). Do children understand fraction addition? *Developmental Science*, *21*(4), e12601. https://doi.org/10.1111/desc.12601

Brousseau, G., Brousseau, N., & Warfield, V. (2004). Rationals and decimals as required in the school curriculum. *The Journal of Mathematical Behavior*, *23*(1), 1–20. https://doi.org/10.1016/j.jmathb.2003.12.001

Charalambous, C. Y., & Pitta-Pantazi, D. (2007). Drawing on a theoretical model to study students' understandings of fractions. *Educational Studies in Mathematics*, *64*(3), 293–316. https://doi.org/10.1007/s10649-006-9036-2

Chen, Z., & Siegler, R. S. (2000). II. Overlapping waves theory. *Monographs of the Society for Research in Child Development*, *65*(2), 7–11. https://doi.org/10.1111/1540-5834.00075

Corwin, R. B., Russell, S. J., & Tierney, C. (1991). *Seeing fractions: A unit for the upper elementary grades*. California Department of Education.

Demedts, F., Kiili, K., Ninaus, M., Lindstedt, A., Reynvoet, B., Sasanguie, D., & Depaepe, F. (2024). The effectiveness of explanatory adaptive feedback within a digital educational game to enhance fraction understanding. *Learning and Instruction*, *94*. https://doi.org/10.1016/j.learninstruc.2024.101976

Demedts, F., Said-Metwaly, S., Kiili, K., Ninaus, M., Lindstedt, A., Reynvoet, B., Sasanguie, D., & Depaepe, F. (2025). Adaptive Feedback in Digital Educational Games: An Explanatory Item Response Theory Approach. *Journal of Computer Assisted Learning*, *41*(5), e70104. https://doi.org/10.1111/jcal.70104

D'Erchie, M., Rosenkranz, J., Schwarzmeier, S., Dresler, T., & Obersteiner, A. (2025). Magnitude-based reasoning can reduce the natural number bias when comparing fractions: An intervention study. *Learning and Instruction*, *98*. https://doi.org/10.1016/j.learninstruc.2025.102109

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

Di Lonardo Burr, S. M., Douglas, H., Vorobeva, M., & Muldner, K. (2020). Refuting misconceptions: Computer tutors for fraction arithmetic. *Journal of Numerical Cognition*, *6*(3), 355–377. https://doi.org/10.5964/jnc.v6i3.310

Di Lonardo Burr, S. M., Xu, C., Li, H., Si, J., LeFevre, J.-A., & Wang, Y. (2022). Fraction mapping and fraction comparison skills among grade 4 Chinese students: An error analysis. *British Journal of Educational Psychology*. https://doi.org/10.1111/bjep.12503

Dowle, M., & Scrinivasan, A. (2021). _*data.table: Extension of 'data.frame'*_ [Computer software]. https://CRAN.R-project .org/package=data.table

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), 1–16. https://doi.org/10.1037/xlm0000153

Gabriel, F., Van Hoof, J., Gómez, D. M., & Van Dooren, W. (2023). Obstacles in the development of the understanding of fractions. In K. M. Robinson, A. K. Dubé, & D. Kotsopoulos (Eds), *Mathematical Cognition and Understanding* (pp. 209–225). Springer International Publishing. https://doi.org/10.1007/978-3-031-29195-1_11

Gómez, D. M., & Dartnell, P. (2019). Middle schoolers' biases and strategies in a fraction comparison task. *International Journal of Science and Mathematics Education*, *17*(6), 1233–1250. https://doi.org/10.1007/s10763-018-9913-z

González-Forte, J. M., Fernández, C., & Van Dooren, W. (2020). Is there a Gap or Congruency Effect? A Cross-Sectional Study in Students' Fraction Comparison. *Studia Psychologica*, *62*(2), 109–122. https://doi.org/10.31577/sp.2020.02.794

González-Forte, J. M., Fernández, C., Van Hoof, J., & Van Dooren, W. (2023). Incorrect ways of thinking about the size of fractions. *International Journal of Science and Mathematics Education*, *21*(7), 2005–2025. https://doi.org/10.1007/s10763-022-10338-7

Hallett, D., Nunes, T., & Bryant, P. (2010). Individual differences in conceptual and procedural knowledge when learning fractions. *Journal of Educational Psychology*, *102*(2), 395–406. https://doi.org/10.1037/a0017486

Halme, H., Van Hoof, J., Hannula-Sormunen, M., & McMullen, J. (2024). Not realizing that you don't know: Fraction state anxiety is reduced by natural number bias. *British Journal of Educational Psychology*, *94*(1), 138–150. https://doi.org/10.1111/bjep.12637

Hansen, N., Jordan, N. C., & Rodrigues, J. (2017). Identifying learning difficulties with fractions: A longitudinal study of student growth from third through sixth grade. *Contemporary Educational Psychology*, *50*, 45–59. https://doi.org/10.1016/j.cedpsych.2015.11.002

Harel, G., & Confrey, J. (1994). *The development of multiplicative reasoning in the learning of mathematics*. State University of New York Press.

Hartig, F. (2024). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. https://CRAN.R-project.org/package=DHARMa

Karamarkovich, S. M., & Rutherford, T. (2019). Fraction errors in a digital mathematics environment: Latent class and transition analysis. *Journal of Numerical Cognition*, *5*(2), 158–188. https://doi.org/10.5964/jnc.v5i2.150

Kieren, T. E. (1976). On the mathematical, cognitive, and instructional foundations of rational numbers. In R. Lesh (Ed.), *Number and measurement: Papers from a research workshop ERIC/SMEAC* (pp. 101–144).

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lamon, S. J. (2020). *Teaching Fractions and Ratios for Understanding: Essential Content 83 and Instructional Strategies for Teachers* (4th edn). Routledge. https://doi.org/10.4324/9781003008057

Lenth, R. V. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans

Lenz, K., Reinhold, F., & Wittmann, G. (2024). Transitions between conceptual and procedural knowledge profiles. Patterns in understanding fractions and indicators for individual differences. *Learning and Individual Differences*, *116*, 102548. https://doi.org/10.1016/j.lindif.2024.102548

Liu, C., Xin, Z., & Li, X. (2012). The development of Chinese students' understanding of the concept of fractions from fifth to eighth grade. *Journal of Mathematics Education*, *5*(1), 45–62.

Mack, N. K. (1995). Confounding whole-number and fraction concepts when building on informal knowledge. *Journal for Research in Mathematics Education*, *26*(5), 422. https://doi.org/10.2307/749431

Malone, A. S., & Fuchs, L. S. (2017). Error patterns in ordering fractions among at-risk fourth-grade students. *Journal of Learning Disabilities*, *50*(3), 337–352. https://doi.org/10.1177/0022219416629647

Maricau, F., Reinhold, F., Leuders, T., Loibl, K., Van Dooren, W., Said-Metwaly, S., Reynvoet, B., & Depaepe, F. (2026). From replication to extension: Unveiling student profiles in fraction comparison tasks. *Contemporary Educational Psychology*, *85*, 102454. https://doi.org/10.1016/j.cedpsych.2026.102454

Marshall, S. P. (1993). Assessment of rational number understanding: A schema-based approach. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds), *Rational Numbers: An Integration of Research* (pp. 261–288). Lawrence Erlbaum Associates.

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. https://doi.org/10.1016/j.learninstruc.2013.12.004

McMullen, J., & Van Hoof, J. (2020). The role of rational number density knowledge in mathematical development. *Learning and Instruction*, *65*, 101228. https://doi.org/10.1016/j.learninstruc.2019.101228

Meert, G., Grégoire, J., & Noël, M.-P. (2010). Comparing the magnitude of two fractions with common components: Which representations are used by 10- and 12-year-olds? *Journal of Experimental Child Psychology*, *107*(3), 244–259. https://doi.org/10.1016/j.jecp.2010.04.008

Metcalfe, J., Xu, J., Vuorre, M., Siegler, R., Wiliam, D., & Bjork, R. A. (2025). Learning from errors versus explicit instruction in preparation for a test that counts. *British Journal of Educational Psychology*, *95*(1), 11–25. https://doi.org/10.1111/bjep.12651

Middleton, J. A., Helding, B., Megowan-Romanowicz, C., Yang, Y., Yanik, B., Kim, A., & Oksuz, C. (2015). A longitudinal study of the development of rational number concepts and strategies in the middle grades. In J. A. Middleton, J. Cai, & S. Hwang (Eds), *Large-Scale Studies in Mathematics Education* (pp. 265–289). Springer International Publishing. https://doi.org/10.1007/978-3-319-07716-1_12

Ni, Y. (1999). Grade 5 and 6 students' understanding of the fraction's meaning. *Psychological Development and Education*, *11*(4), 26–30.

Ni, Y. (2001). Semantic domains of rational numbers and the acquisition of fraction equivalence. *Contemporary Educational Psychology*, *26*(3), 400–417. https://doi.org/10.1006/ceps.2000.1072

Ni, Y., & Zhou, Y.-D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, *40*(1), 27–52. https://doi.org/10.1207/s15326985ep4001_3

Pearn, C., & Stephens, M. (2004). Why you have to probe to discover what year 8 students really think about fractions. In I. Putt, R. Faragher, & M. McLean (Eds), *Mathematics education for the third millenium: Towards 2010 (Proceedings of the 27th Annual Conference of the Mathematics Educa- tion Research Group of Australasia)* (pp. 430–437). MERGA.

R Core Team. (2024). *R* (Version Race for Your Life 4.4.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Reinhold, F., Leuders, T., & Loibl, K. (2023). Disentangling magnitude processing, natural number biases, and benchmarking in fraction comparison tasks: A person-centered Bayesian classification approach. *Contemporary Educational Psychology*, *75*, 102224. https://doi.org/10.1016/j.cedpsych.2023.102224

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. https://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. https://doi.org/10.1037/dev0000275

Saxe, G. B., Taylor, E. V., McIntosh, C., & Gearhart, M. (2005). Representing fractions with standard notation: A developmental analysis. *Journal for Research in Mathematics Education*, *36*(2), 137–157. https://doi.org/https://www.jstor.org/stable/30034828

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. Oxford University Press.

Siegler, R. S., & Lortie-Forgues, H. (2017). Hard lessons: Why rational number arithmetic is so difficult for so many people. *Current Directions in Psychological Science*, *26*(4), 346–351. https://doi.org/10.1177/0963721417700129

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

Sjoberg, D. (2024). *ggsankey: Sankey, Alluvial and Sankey Bump Plots* (Version 0.0.99999) [Computer software]. https://github.com/davidsjoberg/ggsankey

Smith, C. L., Solomon, G. E. A., & Carey, S. (2005). Never getting to zero: Elementary school students' understanding of the infinite divisibility of number and matter. *Cognitive Psychology*, *51*(2), 101–140. https://doi.org/10.1016/j.cogpsych.2005.03.001

Smith, J. P., III, diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, *3*(2), 115–163. https://doi.org/10.1207/s15327809jls0302_1

Spitzer, M. W. H., & Moeller, K. (2022). Predicting fraction and algebra achievements online: A large-scale longitudinal study using data from an online learning environment. *Journal of Computer Assisted Learning*, *38*(6), 1797–1806. https://doi.org/10.1111/jcal.12721

Stafylidou, S., & Vosniadou, S. (2004). The development of students' understanding of the numerical value of fractions. *Learning and Instruction*, *14*(5), 503–518. https://doi.org/https://doi.org/10.1016/j.learninstruc.2004.06.015

Tatsuoka, K. K. (1984). *Analysis of errors in fraction addition and subtraction problems* (Final Report NIE-G-0002; p. 69). Urbana: University of Illinois. https://eric.ed.gov/?id=ED257665

Vamvakoussi, X. (2015). The development of rational number knowledge: Old topic, new insights. *Learning and Instruction*, *37*, 50–55. https://doi.org/10.1016/j.learninstruc.2015.01.002

Vamvakoussi, X., Christou, K. P., Mertens, L., & Van Dooren, W. (2011). What fills the gap between discrete and dense? Greek and Flemish students' understanding of density. *Learning and Instruction*, *21*(5), 676–685. https://doi.org/10.1016/j.learninstruc.2011.03.005

Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2012). Naturally biased? In search for reaction time evidence for a natural number bias in adults. *The Journal of Mathematical Behavior*, *31*(3), 344–355. https://doi.org/10.1016/j.jmathb.2012.02.001

Vamvakoussi, X., & Vosniadou, S. (2010). How many decimals are there between two fractions? Aspects of secondary school students' understanding of rational numbers and their notation. *Cognition and Instruction*, *28*(2), 181–209. https://doi.org/10.1080/07370001003676603

Van Hoof, J., Degrande, T., Ceulemans, E., Verschaffel, L., & Van Dooren, W. (2018). Towards a mathematically more correct understanding of rational numbers: A longitudinal study with upper elementary school learners. *Learning and Individual Differences*, *61*, 99–108. https://doi.org/10.1016/j.lindif.2017.11.010

Van Hoof, J., Degrande, T., Van Dooren, W., & Verschaffel, L. (2017). The transition from natural to rational number knowledge. In D. C. Geary, D. B. Berch, R. Ochsendorf, & K. M. Koepke (Eds), *Acquisition of complex arithmetic skills and higher-order mathematics concepts* (pp. 101–123). Elsevier.

Van Steenbrugge, H., Lesage, E., Valcke, M., & Desoete, A. (2014). Preservice elementary school teachers' knowledge of fractions: A mirror of students' knowledge? *Journal of Curriculum Studies*, *46*(1), 138–161. https://doi.org/10.1080/00220272.2013.839003

VanLehn, K. (1999). Rule-learning events in the acquisition of a complex skill: An evaluation of cascade. *Journal of the Learning Sciences*, *8*(1), 71–125. https://doi.org/10.1207/s15327809jls0801_3

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*(1), 45–69. https://doi.org/10.1016/0959-4752(94)90018-3

Werner, H. (1958). *Comparative psychology of mental development*. Follet Publishing Company.

Wickham, H. (2016). *Ggplot2*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy Messy Data. R package.* (Version 1.3.1) [Computer software]. https://CRAN.R-project.org/package=tidyr

Xu, C., Di Lonardo Burr, S., McNally-Edwards, N., McShane, C., Wan, L., & Wylie, J. (2024). A holistic investigation of fraction learning: Examining the hierarchy of fraction skills, misconceptions, mathematics anxiety and response confidence. *Journal of Cognitive Psychology*, 1–19. https://doi.org/10.1080/20445911.2024.2388907

Zhang, D., & Siegler, R. S. (2022). Curriculum standards and textbook coverage of fractions in high-achieving East-Asian countries and the United States. *Current Opinion in Behavioral Sciences*, *47*, 101184. https://doi.org/10.1016/j.cobeha.2022.101184

Zhang, D., Stecker, P., & Beqiri, K. (2017). Strategies students with and without mathematics disabilities use when estimating fractions on number lines. *Learning Disability Quarterly*, *40*(4), 225–236. https://doi.org/10.1177/0731948717704966

Zhang, Q., & Fiorella, L. (2023). An integrated model of learning from errors. *Educational Psychologist*, *58*(1), 18–34. https://doi.org/10.1080/00461520.2022.2149525

Supplementary Materials

Response Coding and Reliability Procedures

Response coding followed a pre-registered scheme (see OSF) developed prior to data analysis. Raw student responses were independently double-entered and cleaned by the two lead authors[4], with discrepancies resolved by checking against the original paper submissions. This produced a verified baseline dataset used for all subsequent analyses. Two tasks (i.e., fraction number line, unequal partitions) required reasoning in the context of fractions. The interpretation of these responses was also double coded. Because these problems could have multiple valid answers, for approximately 10% of the dataset (n = 100) both researchers independently assessed the responses; inter-rater reliability was very high (Cohen's ks > .90). Any discrepancies were discussed until agreement was reached.

The first two authors then applied deterministic, script-based classification procedures in R to assign responses to predefined strategy and misconception categories across all tasks. To handle cases in which a response could plausibly map onto more than one category, a hierarchical decision rule was implemented in the R script, prioritizing conceptual misconception categories (e.g., whole number bias) over lower-level procedural errors. Responses not initially captured by the predefined coding scheme were systematically screened. Any recurring response pattern observed in at least 20 students or 2% of the sample (whichever was greater) was flagged for joint review by the first two authors to determine whether a strategy could be defined. Where appropriate, these patterns were incorporated into the coding scheme and re-applied to the full

[4] For responses written in Chinese characters, the data were entered by the second author and a trained research assistant, both of whom are native Chinese speakers.

dataset via the R pipeline. Responses that did not meet these criteria were retained as "mixed/uncategorized," and missing responses were coded as "blank."

Developmental Growth and Response Stability

Across all time points, we categorized responses for part-whole, measurement, arithmetic/operator, and density understanding as: i) consistent with whole number bias/applying whole number strategies or procedures to fraction problems (WNB); ii) consistent with partial understanding of fractions (Part); iii) correct (C); iv) mixed or blank (M/B). In Table S1, we specify which responses were considered WNB and which were considered Part for each task and subconstruct. Category labels reflect the labels used in the main text figures. Table S2 presents mean normalized Shannon entropy values by subconstruct, where higher values indicate greater longitudinal variability in response category usage across time points. Table S3 reports Tukey-adjusted pairwise comparisons of entropy derived from the linear mixed-effects model, allowing inference on differences in response stability between subconstructs.

Table S1

Response Categories Across Subconstructs

Subconstruct	Task	WNB Responses	Part Responses
Part-Whole	Fraction Mapping	WNB	S-U; I; Un
Part-Whole	Problem Solving	WNB	S-U; I; Un; MI
Measurement	Fraction Comparison	I-I	I-D
Measurement	Fraction Ordering	I-I	I-D; F<1; D-B; L-S
Measurement	Round to Nearest Whole	WNB; OND	Add
Arithmetic/Operators	Fraction Arithmetic	OND	S-U; I; BD; SD; CD
Arithmetic/Operators	Half of a Half	IF	WO; Add
Density	Between $\frac{1}{4}$ and $\frac{1}{2}$	CF	Op; UF; Seq

Note: Labels are consistent with those from main text figures and Tables S4-S11.

Table S2

Normalized Response Entropy by Subconstruct

Subconstruct	Entropy (SE)	Characterization
Density	.524 (.005)	High inconsistency: Frequent category shifting
Measurement	.365 (.005)	Moderate inconsistency: Frequent conceptual shifting
Arithmetic/Operator	.210 (.005)	Lowest inconsistency: Predominantly stable incorrect trajectories
Part-Whole	.286 (.006)	Low inconsistency: Consistent correct strategy use over time

Note. Values are estimated marginal means from a linear mixed-effects model with subconstruct as predictor and mean task accuracy as covariate.

Table S3

Pairwise Comparisons of Response Entropy

Comparison	Estimate (b)	SE	*z*-ratio	*p*-value
Density vs. Measurement	0.158	0.007	24.14	< .001
Density vs. Arithmetic	0.313	0.007	47.69	< .001
Density vs. Part-Whole	0.238	0.009	27.91	< .001
Measurement vs. Arithmetic	0.155	0.007	23.66	< .001
Measurement vs. Part-Whole	0.080	0.009	9.08	< .001
Arithmetic vs. Part-Whole	-0.075	0.009	-8.48	< .001

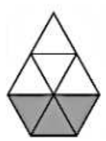
Note. All pairwise comparisons were statistically significant after Tukey adjustment. These results indicate systematic differences in longitudinal response stability across subconstructs.

Qualitative Pathways of Conceptual Change

To address our third research question, we used a response-centred descriptive approach to identify the specific errors and misconceptions that defined trajectories. Here, we present frequency tables capturing responses for each task at each time point (Tables S4-S11). We then provide detailed transition statistics for each task, organized by construct.

Table S4

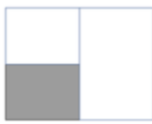
Part-Whole Assessment Responses at Each Time Point: Mapping and Problem Solving

Example Problem	Response Type	Example Responses	Explanation	% Mapping Responses (Modal Responses)				% Pizza/Rice Responses			
				T1	T2	T3	T4	T1	T2	T3	T4
	Correct (C)	$\frac{3}{7};\frac{1}{4};\frac{7}{8}$	Student provided correct response.	69.6 (74.7)	89.2 (94.5)	92.5 (95.5)	95.7 (98.6)	73.9/ 65.6	84.4/ 79.6	85.1/ 84.5	89.2/ 88.4
	Blank (B)	--	Student did not provide a response.	8.8 (5.0)	3.0 (0.8)	1.9 (0.6)	1.0 (0.7)	6.8/ 13.4	4.5/ 7.8	2.7/ 4.1	2.6/ 4.0
	Whole Number Bias (WNB)	3; 1; 4; 7	Student did not consider the ‘part’ and the ‘whole’.	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0/ 3.9	0.0/ 2.2	0.0/ 1.5	0.0/ 0.3
	Part-whole/ Part of a Unit (S-U)	$\frac{3}{4};\frac{1}{3};\frac{7}{1}$	Student responded with shaded portion as numerator and unshaded portion as denominator.	4.2 (4.4)	1.6 (1.4)	1.4 (1.6)	0.5 (0.5)	2.0/ 0.0	1.7/ 0.0	1.1/ 0.0	0.4/ 0.0
	Inverting (I)	$\frac{7}{3};\frac{4}{1};\frac{8}{7}$	Student responded with denominator over numerator.	11.6 (13.0)	1.3 (1.4)	1.3 (1.3)	0.0 (0.0)	3.5/ 3.0	0.4/ 0.6	0.6/ 1.0	0.1/ 1.1
	Unshaded (Un)	$\frac{4}{7}$	Student responded with the unshaded portion as the numerator.	2.6 (2.1)	2.5 (1.0)	1.2 (0.7)	0.8 (0.1)	--	--	--	--
	Misinterpreted (MI)	$\frac{3}{4};\frac{1}{7};\frac{1}{8}$	Student misinterpreted the question (e.g., stating uneaten rather than eaten pieces of pizza, stating number of pieces of rice per day).	--	--	--	--	10.3/ 10.0	6.6/ 7.0	7.8/ 5.4	5.7/ 4.8
	Mixed (M)	--	Response could not be classified or occurred very infrequently.	3.1	2.4	1.7	1.9	3.5/ 4.2	2.4/ 1.3	2.7/ 3.4	2.0/ 1.4

Note. Mapping: N_{T1} = 1,034; N_{T2} = 1,041; N_{T3} = 994; N_{T4} = 1,002. Problem solving: N_{T1} = 1,034; N_{T2} = 1,042; N_{T3} = 992; N_{T4} = 1,001

Table S5

Part-Whole Assessment Responses at Each Time Point: Unequal Partitions

Response	**Explanation**	**Percentage of Responses**			
		T1	T2	T3	T4
Is Sam Correct?					
	Student correctly stated that Sam was not correct.	78.5	85.7	87.3	93.9
	Student incorrectly stated that Sam was correct.	4.2	4.2	4.6	1.6
	Student provided no response.	17.4	10.1	8.1	4.5
Explanations for Response					
No Explanation (None)	Blank response or 'I don't know.'	28.3	17.6	15.2	9.1
Unequal Division (Equal)	Student's explanation mentioned unequal division or the need for equal division.	44.0	50.6	56.6	72.4
Size of Partitions (Size)	Student's explanation made mention of size (i.e., too big, too small, asymmetrical).	9.9	11.6	13.3	5.5
Shapes, Parts, & Lines of Figure (Fig)	Student's explanation referred to shapes (i.e., squares vs. rectangles), parts (i.e., too few), or lines (i.e., missing) in the figure.	5.9	8.0	5.5	4.4
Drawing of Correct Partition (Draw)	Student provided a correct drawing, either of $\frac{1}{3}$ or $\frac{1}{4}$.	5.4	3.3	2.2	3.0
One-Fourth ($\frac{1}{4}$)	Student said that the fraction represented was $\frac{1}{4}$, not $\frac{1}{3}$.	2.3	3.0	2.0	2.6
Incorrect Explanation (I)	Student provided an incorrect explanation.	4.2	6.0	5.1	3.0

Note. N_{T1} = 1,034; N_{T2} = 1,042; N_{T3} = 992; N_{T4} = 1,001

Table S6

Measurement Responses at Each Time Point: Comparison and Ordering

Response Type	Example Response(s)	Explanation	% Comparison Responses				% Ordering Responses $\frac{1}{4},\frac{1}{8},\frac{4}{7},\frac{8}{9}$				% Ordering Responses $\frac{2}{3},1,\frac{5}{4},\frac{4}{9}$			
			T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Correct (C)	$\frac{1}{8},\frac{1}{4},\frac{4}{7},\frac{8}{9};\ \frac{4}{9},\frac{2}{3},1,\frac{5}{4}$	Student provided correct response.	38.0	43.1	50.8	88.9	4.8	6.2	9.9	29.6	2.9	2.8	5.2	34.9
Blank (B)	--	Student did not provide a response.	--	--	--	--	3.1	2.0	1.6	4.5	4.3	3.2	2.5	5.2
Increase-Increase (I-I)	$\frac{1}{4},\frac{1}{8},\frac{4}{7},\frac{8}{9}$	Student compared/ordered the fractions by increasing numerators or increasing denominators.	53.1	47.8	39.2	8.7	49.2	44.0	34.9	16.3	51.1	45.7	37.5	17.5
Increase-Decrease (I-D)	$\frac{8}{9},\frac{4}{7},\frac{1}{8},\frac{1}{4};\frac{8}{9},\frac{1}{8},\frac{4}{7},\frac{1}{4}$	Student compared/ordered the fractions by decreasing numerators or decreasing denominators.	8.6	9.0	10.0	2.4	29.6	32.5	35.0	21.9	30.6	36.6	39.7	19.2
Unit-Big (F<1)	$\frac{4}{9},\frac{2}{3},\frac{5}{4},1$	Student responded with '1' as the biggest fraction, suggesting that the unit is bigger than all fractions.	--	--	--	--	--	--	--	--	2.7	4.8	5.4	5.8
Distance Between (D-B)	$\frac{1}{8},\frac{4}{7},\frac{1}{4},\frac{8}{9};\frac{8}{9},\frac{1}{4},\frac{4}{7},\frac{1}{8}$	Student subtracted numerator from denominator then ordered these differences from smallest to largest or largest to smallest.	--	--	--	--	3.1	4.5	4.9	5.1	--	--	--	--
Largest to Smallest (L-S)	$\frac{8}{9},\frac{4}{7},\frac{1}{4},\frac{1}{8};\frac{5}{4},1,\frac{2}{3},\frac{4}{9}$	Student ordered fractions from largest to smallest.	--	--	--	--	0.0	0.0	0.0	0.0	0.4	0.8	2.2	5.3
Mixed (M)	--	Response could not be classified or occurred very infrequently.	0.0	0.0	0.0	0.0	10.2	10.7	13.7	22.7	8.0	6.2	7.4	12.2

Note. Comparison: N_{T1} = 1,030; N_{T2} = 1,040; N_{T3} = 994; N_{T4} = 1,002; Comparison: N_{T1} = 1,034; N_{T2} = 1,042; N_{T3} = 992; N_{T4} = 1,001

Table S7

Measurement Responses at Each Time Point: Round to the Nearest Whole Number

Response Type	Example Response(s)	Explanation	Percentage of Responses			
			T1	T2	T3	T4
Correct (C)	2	Student provided correct response.	6.4	7.6	10.0	38.0
Blank (B)	--	Student did not provide a response.	18.2	13.6	11.7	14.1
Whole Number Bias (WNB)	40; 19; 21	Student does not consider the relation between the numerator and denominator. They may add numerators, add denominators, add all components, etcetera.	29.1	46.4	41.9	11.4
Operations on Numerator and Denominator (OND)	$\frac{19}{21}$; 1	Student erroneously adds the numerators and denominators, either providing an exact response or rounding to the nearest whole number.	37.9	26.6	32.1	25.7
Add Instead of Estimate (Add)	$\frac{65}{36}$	Student performs fraction arithmetic procedures rather than estimating/rounding.	0.0	0.0	0.0	5.0
Mixed (M)	--	Response could not be classified or occurred very infrequently.	8.4	5.8	4.3	5.9

Note. N_{T1} = 1,034; N_{T2} = 1,042; N_{T3} = 992; N_{T4} = 1,001

Table S8

Arithmetic Responses at Each Time Point: Common Denominator

Response Type	Example Response(s)	Explanation	Percentage of Responses Symbolic								Percentage of Responses Non-Symbolic							
			Addition				Subtraction				Addition				Subtraction			
			T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Correct (C)	$\frac{3}{7}+\frac{2}{7}=\frac{5}{7}$	Student provided correct response.	48.6	53.1	72.1	90.8	54.6	63.9	80.4	91.7	58.3	66.1	68.5	86.9	60.6	65.9	68.5	84.4
Blank (B)	--	Student did not provide a response.	18.4	10.1	7.3	5.2	22.0	11.7	7.7	5.4	0.8	0.1	0.5	0.5	2.6	1.6	0.7	0.9
Part-whole/ Part of a Unit (S-U)	$\frac{3}{7}+\frac{2}{7}=\frac{5}{9}$	Student added/subtracted the shaded pieces for the numerator and added/subtracted the unshaded pieces for the denominator.	--	--	--	--	--	--	--	--	4.5	2.7	1.5	0.5	3.6	1.6	1.7	0.2
Inverting (I)	$\frac{3}{7}+\frac{2}{7}=\frac{7}{5}$	Student applied the correct arithmetic procedure but inverted the final response.	--	--	--	--	--	--	--	--	2.8	0.4	0.3	0.0	2.7	0.2	0.3	0.2
Operations on Numerator and Denominator (OND)	$\frac{3}{7}+\frac{2}{7}=\frac{5}{14}$	Student performed operations on both the numerator and denominator (i.e., adding/subtracting both numerators and denominators).	28.0	32.9	17.4	2.0	19.5	20.0	8.7	1.2	23.7	24.3	20.9	4.4	19.2	21.4	19.6	5.7
Mixed (M)	--	Response could not be classified or occurred very infrequently.	4.9	3.9	3.2	2.0	3.9	4.4	3.2	1.7	9.8	6.4	8.3	7.7	11.3	9.2	9.2	8.6

Note. N_{T1} = 1,034; N_{T2} = 1,042; N_{T3} = 992; N_{T4} = 1,001

Table S9

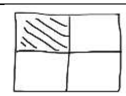
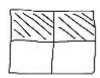
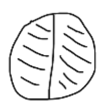
Arithmetic Responses at Each Time Point: Different Denominators

Response Type	Example Response(s)	Explanation	Percentage of Responses Symbolic								Percentage of Responses Non-Symbolic							
			Addition				Subtraction				Addition				Subtraction			
			T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Correct (C)	$\frac{1}{2}+\frac{2}{3}=\frac{7}{6}$	Student provided correct response.	0.1	0.1	0.7	44.2	0.0	0.0	0.7	42.5	0.2	0.6	0.6	22.1	0.1	0.4	0.5	17.4
Blank (B)	--	Student did not provide a response.	21.4	11.5	9.0	6.7	38.9	21.5	17.0	11.9	7.8	5.3	3.6	5.4	5.8	4.2	3.4	4.2
Operations on Numerator and Denominator (OND)	$\frac{1}{2}+\frac{2}{3}=\frac{3}{5}$	Student performs operations on both the numerator and denominator (i.e., adding/subtracting both numerators and denominators).	68.9	74.1	74.7	31.0	36.2	38.7	40.7	11.9	46.4	44.7	52.9	20.7	28.6	25.4	27.7	12.1
Big Denominator (BD)	$\frac{1}{2}+\frac{2}{3}=\frac{3}{3}$	Student added/subtracted the numerators and then selected the larger denominator.	4.9	6.2	7.7	2.3	4.7	10.5	13.9	6.0	27.6	30.6	25.6	24.4	43.9	43.3	42.0	39.3
Small Denominator (SD)	$\frac{1}{2}+\frac{2}{3}=\frac{3}{2}$	Student added/subtracted the numerators and then selected the smaller denominator.	0.7	1.5	0.8	0.7	9.5	16.0	15.6	6.3	6.9	7.0	4.3	3.6	12.8	16.4	13.6	13.3
Common Denominator (CD)	$\frac{1}{2}+\frac{2}{3}=\frac{3}{6}$	Student found a common denominator but did not adjust the numerators correctly.	1.5	0.8	0.7	11.8	0.0	0.3	0.4	11.9	0.4	0.2	0.3	13.1	0.1	0.2	0.2	4.8
Mixed (M)	--	Response could not be classified or occurred very infrequently.	2.4	5.8	6.5	3.4	10.7	13.1	11.6	9.6	10.7	11.6	12.6	10.8	8.7	10.1	12.5	9.0

Note. N_{T1} = 1,034; N_{T2} = 1,042; N_{T3} = 992; N_{T4} = 1,001

Table S10

Operator Responses at Each Time Point: Half of a Half

Response Type	Example Response(s)	Explanation	Percentage of Responses			
			T1	T2	T3	T4
Correct (C)		Student provided correct response.	29.2	36.5	45.2	69.3
Blank (B)	--	Student did not provide a response.	6.0	4.9	3.2	1.9
Independent Fractions (IF)		Student thinks about the two halves as being independent.	53.4	45.6	41.4	22.2
Whole Only (WO)		Student understands that they need to partition the shape in half and then in half again, but they do not understand how that links to quantity (i.e., they know the whole but not the part).	1.4	3.1	2.2	2.1
Added Two Halves (Add)		Student believes they are to add two halves together.	2.7	2.3	2.4	0.9
Mixed (M)	--	Response could not be classified or occurred very infrequently.	7.4	7.7	5.5	3.6

Note. N_{T1} = 1,034; N_{T2} = 1,042; N_{T3} = 992; N_{T4} = 1,001

Table S11

Density Responses at Each Time Point: Fractions Between $\frac{1}{4}$ and $\frac{1}{2}$

Response Type	Example Response(s)	Explanation	Percentage of Responses			
			T1	T2	T3	T4
Correct (C)	Infinite	Student provided correct response.	15.4	31.5	38.2	67.8
Blank (B)	--	Student did not provide a response.	24.4	12.7	10.4	6.6
Countable Fractions (CF)	Whole numbers (e.g., between 2 and 20)	Student believes that there is a finite, countable number of fractions.	29.7	33.7	32.7	8.9
Performed an Operation (Op)	$\frac{1}{2}; \frac{0}{2}; \frac{0}{4}$	Student (often erroneously) performed an operation on the two fractions, such as adding or subtracting.	14.1	6.4	6.5	3.5
Unit Fractions Only (UF)	$\frac{1}{3}$; 1	Student believes that fractions increase with respect to unit fractions (e.g., $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$ ...) and that no other fractions exist between these fractions.	8.9	10.7	7.5	7.8
Sequence of Fractions (Seq)	$\frac{1}{1}, \frac{1}{2}, \frac{2}{1}, \frac{2}{2}$	Student responded with a sequence of fractions, trying to list all fractions (often erroneously) between the two fractions.	7.4	5.0	4.8	5.4

Note. N_{T1} = 1,034; N_{T2} = 1,042; N_{T3} = 992; N_{T4} = 1,001

Early-Stabilizing Concepts: Part-Whole Understanding

Fraction Mapping. As shown in Table S4, accuracy on fraction mapping items was high at T1 and reached near-ceiling by T4. Beyond these aggregate distributions, the longitudinal data reveals a high degree of stability once mastery was achieved. After the time point in which students first provided a correct response, 96.0% of students remained in the correct category for all subsequent time points. Incorrect responses were infrequent following the initial assessment; only 0.7% of students failed to transition into the correct response category at any point during the study.

Problem Solving. While accuracy for the pizza and rice problems was high across all time points (see Table S4), individual pathways were highly diverse, with 117 and 137 distinct transition patterns identified, respectively. Despite this individual variety, students rarely remained stuck in a specific misconception. Excluding students who were correct at all four time points, only 0.1% (pizza) and 0.2% (rice) of students maintained the same incorrect response category throughout the study. Furthermore, after being first classified as correct, three-quarters of students (73% pizza, 75% rice) maintained that mastery for all subsequent time points. Overall, only 1-2 students in the entire sample failed to provide a correct response at any of the four time points.

Unequal Partitions. Analysis of the rationales provided for the unequal partitions task (Table S5) highlights a clear developmental shift toward the unequal division explanation. While only 51 transition patterns existed for the *"Is Sam correct?"* prompt, the rationales included 339 distinct transition patterns, reflecting the complexity of student reasoning. Individual-level shifts were primarily directional toward mastery rather than being characterized by random fluctuations. From T1 to T2, 10% of students moved from providing no explanation to the unequal division category. This stabilization continued into later time points: the proportion of students remaining in the unequal division category grew from 37% (T2-T3) to 49% (T3-T4). Across time points, students were more likely to move from a blank, incorrect, or partially correct explanation to a correct one (19% to 23% per time point) than to revert to an incorrect explanation (7% to 13%). By T4, 56% of students showed stable membership in the correct rationale category.

Transitional Misconceptions: Measurement and Magnitude Understanding

Fraction Comparison. As shown in Table S6, mean scores for fraction comparison increased from 13.9% at T1 to 18.2% at T4. However, these means were heavily driven by performance on the 14 congruent trials, where larger components align with larger magnitude. Median scores across these congruent trials remained high (ranging from 11 to 13), while median scores for the 9 incongruent trials were near-zero (0, 1, and 4) for the first three time points before rising to 8 at T4. This discrepancy indicates that early in the study, most students' responses aligned with whole number bias. A strategy-switch phenomenon was observed at T3, where over 100 students scored 0 on congruent trials despite success on incongruent trials, suggesting they had incorrectly over-generalized that larger components always signify a smaller fraction. Across the study, more students switched from a whole-number-bias strategy (i.e., increase-increase) to this increase-decrease strategy (14%) than vice versa (7%). By T4, 89.8% of students were correct on congruent trials and 78.3% were correct on incongruent trials.

Fraction Ordering. Ordering four fractions was more difficult for students than pairwise comparison, with accuracy remaining below 10% until T4. Individual pathways were characterized by high volatility, with 284 and 311 distinct transition patterns identified for the two ordering items. Very few students remained in the same category across all time points (8%

and 9% for Items 1 and 2, respectively), and less than 1% provided correct responses at all four time points. This subconstruct also showed high rates of regression; of the students who reached mastery prior to T4, 41% for the Item 1 and 46% for Item 2 reverted to incorrect responses at subsequent time points. These patterns of volatility are visualized in the transition flows in Figure 4 in the main text.

Rounding to the Nearest Whole Number. Transition patterns for the rounding task involved 300 distinct pathways. From T1 to T2, the most common shifts involved remaining in whole-number bias (WNB; 17%) or moving from the operations on numerator and denominator (OND) category back to WNB (17%). During this period, students were just as likely to move from an incorrect to a correct response (6%) as they were to revert (5%). From T2 to T3, roughly 40% of students who were incorrect made consistent errors, with the most common movement being shifts between the WNB and OND categories. However, a significant shift occurred between T3 and T4, as students were much more likely to move from an incorrect to a correct response (32%) than to revert (3%). During this final transition, 14% of students moved from WNB to correct, and 12% moved from OND to the correct category.

Entrenched Misconceptions: Fraction Arithmetic and Operators

Common Denominators. By T4, mastery reached > 90% for symbolic and > 80% for non-symbolic items (see Table S8). The slightly lower early performance on symbolic tasks was largely due to a higher proportion of blank responses, likely because these items appeared at the end of the timed assessment. Transition analyses revealed 102 patterns for symbolic addition, 91 for symbolic subtraction, 122 for non-symbolic addition, and 136 for non-symbolic subtraction. Longitudinally, students commonly remained in the OND category across subsequent time points (21% for symbolic addition, 10% for symbolic subtraction, 18% for non-symbolic addition, and

13% for non-symbolic subtraction). Alternatively, they moved from OND to a correct response (42% for symbolic addition, 27% for symbolic subtraction, 35% for non-symbolic addition, and 31% for non-symbolic subtraction).

Different Denominators. Students struggled with different denominator arithmetic (see Table S9); for all four items, only a single student in the entire sample provided a correct response at all four time points (for non-symbolic addition). Complexity was reflected in the high number of transition patterns: 195 for symbolic addition, 390 for symbolic subtraction, 324 for non-symbolic addition, and 339 for non-symbolic subtraction. By T4, fewer than 45% of students were correct on symbolic arithmetic problems, and fewer than 25% were correct on non-symbolic versions.

Consistency Across Representations and Operations. A central finding in this subconstruct was the inconsistency of student strategies. When students made an error on common denominator items, the vast majority used different error categories for symbolic versus non-symbolic problems (ranging from 75% to 97% across time points). A similar lack of consistency was found for different denominator items, where 46% to 88% of students used different error types across representations.

Comparisons across operations revealed that students frequently shifted their strategy depending on whether they were adding or subtracting. Across the four time points, for non-symbolic arithmetic, the percentage of students using different error categories for addition versus subtraction increased from 53% to 75% for common denominators, and from 50% to 64% for different denominators. For symbolic items, these differences ranged from 42% to 53% for common denominators and 42% to 67% for different denominators. This lack of cross-operation consistency was often driven by the constraints of the numbers; while students frequently applied whole-number logic (i.e., OND) to addition, they were more likely to switch to a selection strategy (i.e., choosing the bigger or smaller denominator) for subtraction to avoid zero or negative results. Overall, these patterns indicate that students' approaches to fraction arithmetic were highly inconsistent across representation and operation.

The Operator Concept (Half of a Half). The independent fractions error accounted for 72% to 76% of all errors at each time point. Across 189 transition patterns, the most common transitions involved students either remaining in the independent fractions category or moving from that category to a correct response. The consistency of this misconception was notable: of the students who were incorrect at any point, between 20% and 37% made consistent errors across time points. However, the move toward mastery became more pronounced by T4, where students were five times more likely to move from an incorrect to a correct response (30%) than to revert to an error (6%).

Conceptual Thresholds: Density Understanding

As shown in Table S11 and Figure 8 (main text), understanding of fraction density reached a threshold at T4 (68% correct). Transition analysis revealed 356 patterns. From T1 to T2, the most common patterns were remaining in the countable category (11%) or moving from countable to correct (9%), with 23% of incorrect students making consistent errors. This stability in error continued from T2 to T3, where 15% remained in the countable category and 27% of incorrect students maintained consistent errors. However, a developmental shift occurred between T3 and T4; the most frequent patterns were remaining in the correct category (31%) and moving from the countable category to correct (21%). During this final transition, the rate of moving from incorrect to correct (37%) was more than five times higher than the rate of reversion (7%), and error consistency dropped to only 12%.

Number Line

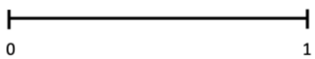
The number line item stated, "*When Ben was asked to place $\frac{4}{7}$ on the number line below, he said that $\frac{4}{7}$ could not be placed on the line because both 4 and 7 are greater than 1. Why is Ben incorrect? Where should he place $\frac{4}{7}$?*" A student's placement was considered correct if it was closer to the fourth segment (when the line is divided into sevenths) than any other integer from 0 to 7, excluding placements at $\frac{1}{2}$. While most students understood that the fraction could indeed be placed on the line, they struggled to provide adequate rationales. A large proportion of students (ranging from 34% to 40% across time points) attempted to place the fraction but provided no explanation (see Table S12).

When explanations were provided, the most frequent appropriate rationale involved dividing the line into seven equal parts or recognizing the endpoint '1' is equal to '$\frac{7}{7}$'. At the earlier time points, students often noted that fractions are not the same as whole numbers or that numerators and denominators cannot be separated. Some students gave partially correct explanations that alluded to all fractions being smaller than 1. At T4, after being introduced to improper fractions, 18% of students specified that $\frac{4}{7}$ was a proper fraction and all proper fractions are smaller than 1. Whole number bias (evidenced by students stating that the right endpoint '1' represented '10') and incorrect decimal interpretations were less common but still present at all four time points. Despite providing appropriate explanations, few students demonstrated precise magnitude knowledge. Incorrect placements were distributed across the line, and mastery was notably unstable. Although 69% of students placed $\frac{4}{7}$ correctly at least once, 43% of these students transitioned back to an incorrect or blank response at a subsequent time point, and only 2% provided a correct placement at all four time points.

Figure S1 illustrates the volatility of this task, which was characterized by 75 different transition patterns for number line placement and 520 patterns for explanations. Only 8% of students provided an appropriate explanation at all four time points, while 25% provided no explanation or an explanation consistent with a misconception throughout the study. From T1 to T2, the most frequent pattern was a consistent lack of explanation (34%), and only 14% of students who provided an explanation remained in the same category. Movement toward an appropriate explanation (23%) outpaced reversion (13%) during this interval. From T2 to T3, students were equally likely to move toward an appropriate explanation (18%) as they were to revert (17%), with only 20% of students maintaining their explanation category. Between T3 and T4, 9% of students consistently stated the line should be divided and 7% moved from no explanation to that rationale. During this final transition, 18% of students remained in the same category, and movement toward an appropriate explanation (24%) again outpaced reversion (15%).

Table S12

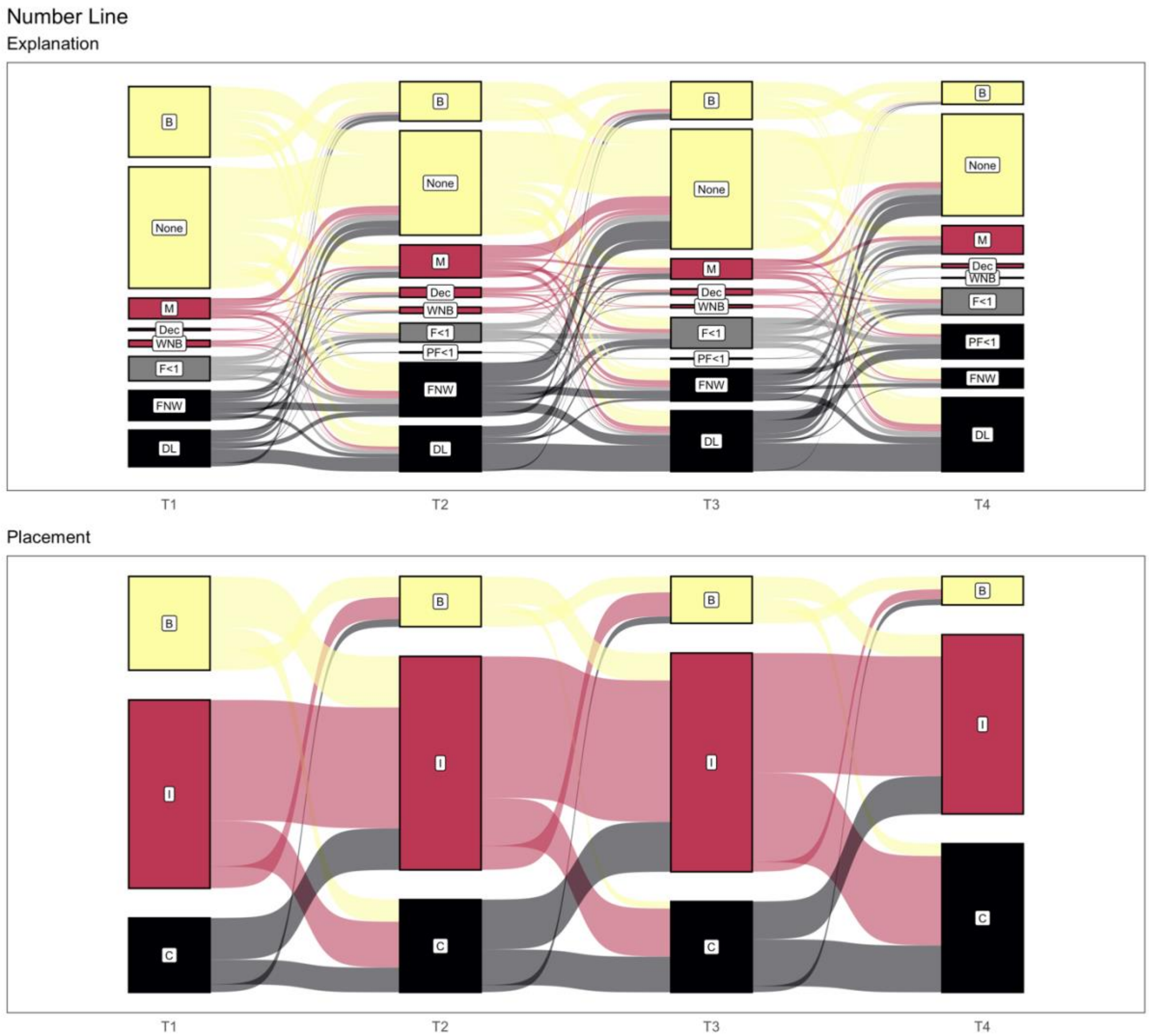
Number Line Responses at Each Time Point

Response	Explanation	Percentage of Responses			
		T1	T2	T3	T4
Number Line Placement					
	[a]Student placed $\frac{4}{7}$ in approximately correct location.	21.9	25.9	25.8	42.0
	Student placed $\frac{4}{7}$ in an incorrect location.	51.8	60.4	61.3	49.8
	Student provided no response.	26.3	13.7	12.9	8.3
Explanations for Response					
Blank (B)	Student did not provide an explanation, nor did they estimate the fraction location on the number line.	22.5	7.6	9.4	6.4
No Explanation (None)	Student did not provide an explanation but did estimate the fraction location on the number line.	39.7	37.9	41.3	33.5
Divide the Line (DL)	Student's explanation mentioned needing to divide the line into 7 parts or thinking of the endpoint "1" as "$\frac{7}{7}$".	11.3	15.1	19.6	23.7
Fractions Not Whole Numbers (FNW)	Student's explanation mentioned fractions being different from whole numbers (e.g., can't separate numerator and denominator).	9.3	16.9	10.2	6.1
Fractions Less Than 1 (F<1)	Student's explanation mentioned (all) fractions being smaller than 1 or fractions being smaller than whole numbers.	7.4	6.4	9.8	8.6
Whole Number Bias (WNB)	Student's explanation mentioned that the endpoint "1" represents "10".	1.9	2.1	1.2	0.2
Decimal (Dec)	Student's explanation mentioned an incorrect conversion from fractions to decimals (e.g., $\frac{4}{7} = 7.4$ or $\frac{4}{7} = 0.47$)	0.7	3.1	1.9	1.2
Proper Fractions Less Than 1 (PF<1)	Student provided an incorrect explanation.	0.0	0.3	0.3	11.1
Mixed (M)	Student provided an incorrect explanation that could not be classified or occurred very infrequently.	7.3	10.7	6.4	9.3

Note. N_{T1} = 1,034; N_{T2} = 1,042; N_{T3} = 992; N_{T4} = 1,001; [a]Correct placements were those where, when dividing the line into seven equal segments, the fraction was placed closer to 4 than any other integer from 0 to 7, and the fraction was not placed at $\frac{1}{2}$.

Figure S1

Proportion and Transition in Category of Response for Number Line



Note. N = 950. C = Correct; I = Incorrect; B = Blank; DL = Divide the line; FNW = Fractions are not whole numbers; F<1 = Fractions are less than 1; WNB = Whole number bias; Dec = Decimals; PF<1 = Proper fractions are less than 1; M = Mixed; None = No explanation provided.