COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 50 (2026) e70168
© 2026 Cognitive Science Society LLC.
ISSN: 1551-6709 online
DOI: 10.1111/cogs.70168

Enhancing Flexible Transfer of Fractions: The Role of Sequential and Simultaneous Games With Multiple Representations

Kreshnik N. Begolli, Siling Guo, Lourdes M. Acevedo-Farag, Giovanni Sanchez, Xiangqian Yu, Yoori Kim, Milan Vu, Laura Hernandez, June Ahn, Drew Bailey, Andres S. Bustamante, Katherine Rhodes, Lindsey Richland

School of Education, University of California, Irvine

Received 14 November 2024; received in revised form 16 November 2025; accepted 5 December 2025

Abstract

We codesigned and evaluated a brief intervention combining two fraction games: Fraction Ball (played on a basketball court) and Bottle Caps Bonanza (played on a tabletop shuffleboard). Using participatory design principles, we engaged teachers and students in codesigning playful learning experiences aimed at improving knowledge transfer and adding fractions with unlike denominators. Students were randomly assigned within seven treatment classrooms to practice fractions with different denominators on one board simultaneously ($N = 87$) versus practicing on separate boards sequentially ($N = 79$). Three comparison classrooms ($N = 75$) only took the pretest and posttest. Our preregistered models suggested significant impacts on multiple aspects of fraction knowledge, including far transfer and overall fraction knowledge, when comparing both treatment groups to the comparison group. The simultaneous condition performed higher on untimed fraction addition with unequal denominators, though this difference was not statistically significant ($b = 0.21$, $p = .05$). Furthermore, students with higher prior knowledge benefited more from the simultaneous condition. We conclude that this playful and accessible intervention can effectively improve students' fraction knowledge.

Keywords: Playful learning; Knowledge transfer; Fractions; Analogy; Math cognition

[Correction added on 12 February 2026, after first online publication: The spell error in author name 'Laura Hernandez' has been corrected.]

Kreshnik N. Begolli and Siling Guo contributed equally to this manuscript and share first authorship.

Correspondence should be sent to Kreshnik N. Begolli, School of Education, University of California, Irvine, 401 E. Peltason Drive, Suite 3200, Irvine, CA 92617, USA. E-mail: kbegolli@uci.edu

1. Introduction

Play is a key that unlocks children's potential for learning in a meaningful context that is actively engaging and socially interactive (Bustamante et al., 2020; Hassinger-Das, Bustamante, Hirsh-Pasek, & Golinkoff, 2018; Schlesinger et al., 2020). Play contexts offer rich opportunities for promoting knowledge transfer—the cognitive ability to generalize knowledge from the original learning conditions and representations and apply them to new contexts or representations (see Booth et al., 2017; Bransford & Schwartz, 1999; Koedinger, Corbett, & Perfetti, 2012; Richland & Begolli, 2016; Star, Rittle-Johnson, & Durkin, 2016; Zosh et al., 2018), which has been a long standing goal for instruction (Common Core State Standards Initiative, 2012; National Mathematics Advisory Panel, 2008; National Research Council, 2001). Play can also serve as a powerful context for learning challenging mathematical topics, such as fractions (Bustamante, Begolli, Alvarez-Vargas, Bailey, & Richland, 2022), that are important in many professional domains (Handel, 2016) and for advancing to higher mathematical concepts (Booth & Newton, 2012).

Learning math through play in a way that demonstrates improvement on math assessments inherently requires transfer in order to apply knowledge learned in a game context to decontextualized, abstract contexts and formats and/or word problems. Based on the logic of transfer appropriate processing (Schwartz & Bransford, 2005; Morris, Bransford & Franks, 1977), learning through play which incorporates science of learning principles to promote practice with generalization should also lead to children's development of flexible, portable knowledge that can be invoked across contexts (e.g., from a basketball game to a math problem on paper) and number representations (e.g., from fourths to fifths). Generalization could be enhanced when learning across two game contexts that vary on the surface but share the same mathematical concepts (e.g., Gick & Holyoak, 1983).

In this study, we offer an innovative approach for teaching fractions and promoting knowledge transfer through a combination of two games—*Fraction Ball* and *Fraction Bottle Caps Bonanza*. The games were designed through a participatory design process in which our research team, teachers, and students created and iterated on our intervention in collaborative codesign sessions. Through codesign, we integrated research from the science of learning and community expertise to effectively address users' needs, challenges, and expectations, and to ensure the games are indeed fun, to fulfill our goal of learning through play. *Fraction Ball* is a version of basketball in which the three-point line is replaced by a series of arcs labeled with both fraction and decimal representations (see Begolli et al., 2024; Bustamante et al., 2022). A previous efficacy experiment of Fraction Ball generated positive impacts on students' overall rational number knowledge, magnitude estimation on a number line, fraction/decimal conversions, and adding fractions with decimals compared to "business as usual" math instruction and physical education (PE) classes (Begolli et al., 2024; Bustamante et al., 2022). Yet, impacts on fraction values not represented on the court, and on concepts like adding unlike denominators, have not yet been demonstrated (Begolli et al., 2024; Bustamante et al., 2022), suggesting opportunities to strengthen and extend our game impacts.

Through collaboration with students and educators, we aimed to magnify our games' impacts by drawing on codesign to produce playful activities that lead students to grapple

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

with additional rational number concepts. We codesigned with school partners to address these additional learning outcomes, which led to changes in Fraction Ball's gameplay as well as adding a second game, *Bottle Caps Bonanza*, in which students play a modified tabletop version of shuffleboard (Table 1; also see Bustamante et al., 2022). Excitingly, as we will describe through a design narrative (Hoadley, 2023), the design ideas of school partners aligned with theory on learning by analogy, enabling us to design an experiment in which we tested the impact of integrating opportunities for comparing and contrasting representations into the two rational number games.

We report an investigation in which we evaluated the effectiveness of these two new games compared to "business as usual" math and PE class using a quasi-experimental design. Further, we compared two distinct iterations of the games to each other in a randomized A/B experiment to examine the order of presenting comparisons, one in which unlike denominators were presented simultaneously (side-by-side) and a second where they were presented sequentially (one at a time). This is a phenomenon that was of teacher interest and has yielded mixed results in the analogical reasoning literature, which may be because the benefits of order may depend on reasoners' prior knowledge (Rittle-Johnson et al., 2015). For example, in learning about ratios and proportions, presenting solution methods simultaneously improved overall students' conceptual understanding of ratio concepts, while sequential presentations led to increased use of a common misconception (Begolli & Richland, 2016). In algebra, sequential presentations of two solution methods were shown to be more beneficial for low prior knowledge students, whereas simultaneous presentations were more beneficial for high prior knowledge students (Rittle-Johnson, Star, & Durkin, 2009).

Below, we discuss the potential for playful learning to promote far transfer in the context of fraction addition with unlike denominators, as well as to build theory on the relations among play, analogy, and far transfer more broadly. Additionally, we review how the science of learning informed the designs of Bottle Caps Bonanza and modifications of the Fraction Ball game. Then, we highlight how playful designs can be iteratively strengthened and expanded by drawing on the science of learning principles and participatory design principles with school partners.

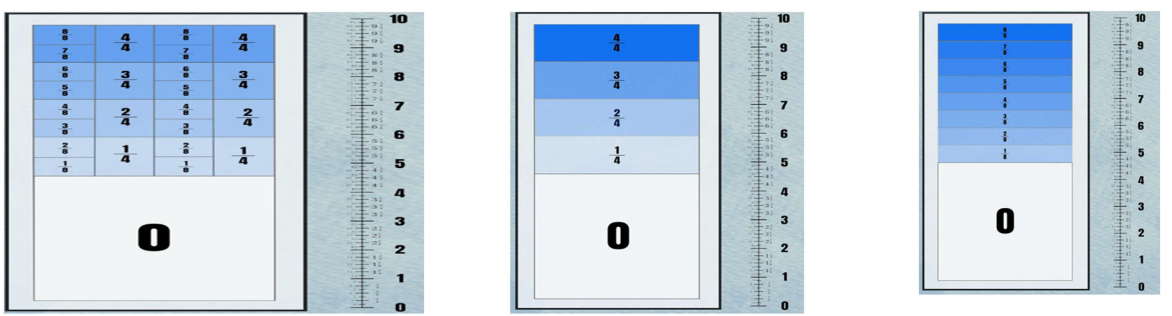
1.1. Playful learning experiences as opportunities for cross- and within-domain transfer

First, we discuss the definition of play, and draw attention to its role as an opportunity to represent content—in this case, mathematics—in two domains: physically grounded game play and decontextualized mathematics problem-solving. We draw from a definition of playful learning that encompasses five key characteristics: actively engaging, joyful, iterative, socially interactive, and meaningful (Hirsh-Pasek et al., 2015). Playful learning entails being actively engaged and attentive to problem-solving using mathematical principles and ideas in service of the goals of play (Authors et al., 2022), which are not solely embedded in symbolic calculations on paper or computer. Theoretically, therefore, play provides practice in transfer in addition to practice with the specific skills, due to its cross-domain nature, allowing for comparing representations of the same mathematical concepts. In our study, children gain practice with adding fractions by playing a basketball game where they have to sum their

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
Redesigned Fraction Ball court and Bottle Caps Bonanza boards

Court type	Simultaneously visible court (fourths and eighths)	Sequentially visible court (left: fourths; right: eighths)
Photo		

Board type	Simultaneously visible board (fourths and eighths)	Sequentially visible board (left: fourths; right: eighths)
Photo		

Note. In *Fraction Ball*, when a shooter scores a shot from the 2/4 area of the court, the rebounder announces the corresponding value, "2/4," and the counter moves forward on the number line, adding 2/4 to their team's score. In *Bottle Caps Bonanz*a, students flick a bottle cap to labeled fraction increments on a tabletop fraction board and add their "flicks" to determine the winning team with the highest score. Fraction Ball courts and Bottle Caps Bonanza boards presenting fourths and eighths (as shown) were used in the first court and in-class activity, respectively. In the second activity, the same designs were used to present thirds and sixths. To modify the representations on the painted court and create our two conditions, sequentially and simultaneously visible, we used blank and printed, spot markers. The top left photo shows $\frac{1}{4}$ painted and a spot marker with 2/8 to show the $\frac{1}{4}$ and 2/8 simultaneously by covering 0.25, which was painted on the court. The same was done for 2/4 and 0.50, and so on. The top middle photo shows a blank spot marker covering the decimals to show fourths only. The top right photo shows a spot marker with 2/8 covering the decimals with fourths covered with blank spot markers to show only eighths. Both the top middle and right are designed for the sequential condition.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

scores in fractional points after each shot. We theorize that transferring the same knowledge across contexts (e.g., adding 1/4s on a number line in two games to adding 1/4s symbolically on a classroom-based paper worksheet) should also lead to flexible, portable knowledge representations. Further, game play that allows for within-domain comparisons, moving between mathematical representations (e.g., comparing fourths and eighths) can also increase flexible representations (e.g., see Alfieri, Nokes-Malach, & Schunn, 2013, Begolli & Richland, 2016, Gray & Holyoak, 2021).

In both cross- and within-domain comparison, children are exposed to experiences that prompt reasoners to deliberate on what are irrelevant (surface) features and core (relevant) features across- and within-domains (Gick & Holyoak, 1983: Gentner, 1983). By aligning the relevant elements in each representation/context and mapping between them, they engage in schema building that promotes generalized knowledge transfer to representations and contexts, unlike those practiced during the play context (e.g., from fourths in two games to tenths on paper worksheets; see Schwartz & Bransford, 2005; Morris, Bransford & Franks, 1977).

This provides theoretical insight into the role of play as a powerful vehicle for building flexible representations of symbolic content. Play may lead to successful future transferable knowledge both through its nature as facilitating connections between cross-domain representations, and due to its ability to incorporate comparison activities naturally within play. We emphasize this point because children often struggle to transfer their skills between modalities or concepts (e.g., see Gray & Holyoak, 2021; Moser et al., 2015). However, we posit that well-constructed play activities can help students gain success with these skills.

In the current study, we build on these principles and extend them to offer learning opportunities in two contexts (Bottle Cap Bonanza and Fraction Ball) with multiple fraction representations (fourths, eighths, thirds, sixths) aligned with the ideas of transfer appropriate processing (Morris, Bransford & Franks, 1977). This provides more opportunities for transferring mathematical domain concepts during the acts of learning by engaging learning processes both across domains and within domains (i.e., across mathematical magnitude representations). In turn, transfer appropriate processing should lead to increasingly more successful problem-solving across contexts (i.e., farther transfer). For example, children engaged in learning about fractions in two different game contexts (basketball and shuffleboard) as well as rational number representations (pairing thirds with sixths and fourths with eighths). Further, Fraction Ball and Bottle Cap Bonanza meet the five characteristics of play, as they are socially interactive (e.g., children play in teams on the court or in-class), actively engaging (observations and teacher reports support this), joyful and meaningful (students report being less bored and more hopeful; Guo et al., 2024). As such, our study provides an example of how to tackle the challenge of utilizing principles from playful learning and analogy for enhancing the transfer of mathematical knowledge, particularly rational number understanding.

1.2. Defining near and far transfer

In this study, we draw from a mathematical perspective to define our notions of near and far transfer. We define near transfer as the ability to problem solve with the same mathematical magnitudes (e.g., $\frac{1}{4}$) across two different contexts (e.g., play vs. decontextualized

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

symbolic problem-solving). For example, this might include engaging with fraction addition with fourths within the game of Fraction Ball versus on a computerized assessment with traditional number symbols and operands. In contrast, we define far transfer as the ability to both transfer across contexts and solve problems that involve mathematical representations not previously encountered during the playful learning experience (e.g., experience only fourths during play but then assessed on addition with fifths, which were not shown during the game). A complete list of problem types is available in the Supplementary Materials, Section A, Table A1. As discussed in more detail below and in the Method section, children manipulated varied rational numbers and operations both during play and on pre/posttest measures. Thus, near (across context) and far transfer items (across context and math representations) included addition of like and unlike denominators, mixed numbers, fractions larger than one, as well as estimation on a 0–1 and 0–5 number lines with and without hatch marks.

1.3. Integrating science of learning principles and play to promote far transfer

As shown in Table 1, the designs of the current Bottle Caps Bonanza and Fraction Ball games drew heavily from the literature on fraction magnitude understanding, comparing multiple representations, and common misconceptions students face with fraction operations. The designs continue to build on the mental representation of the number line, postulated to be a key mental fraction magnitude representation in mathematics (Braithwaite & Siegler, 2021), as supported by positive impacts found in prior intervention studies (Begolli et al., 2024; Bustamante et al., 2022; Barbieri, Rodrigues, Dyson, & Jordan, 2020; Bottge et al., 2014; Dyson, Jordan, Rodrigues, Barbieri, & Rinne, 2020; Fazio, Kennedy, & Siegler, 2016; Fischer, Moeller, Bientzle, Cress, & Nuerk, 2011; Fuchs et al., 2013; Hamdan & Gunderson, 2017; Link, Moeller, Huber, Fischer, & Nuerk, 2013; Schumacher et al., 2018; Siegler & Ramani, 2008).

The number line represented on the Fraction Ball and Bottle Cap Bonanza courts and scoreboards also have implications for addressing fraction misconceptions in several ways. For example, children can "see" that adding $\frac{1}{4}$ and $\frac{1}{4}$ on the number line makes 2/4 (not 2/8); addressing common calculation errors, likely resulting from misapplying whole number procedures, known as whole number bias (Ni & Zhou, 2005). Further, children may recognize the relations between the numerator and denominator as 1 part of 4 (see Dewolf, Bassok, & Holyoak, 2015), representing a single unit that can be concatenated (e.g., $\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}$) to get 1 whole (Braithwaite & Siegler, 2021). As such, the board and court aim to strengthen children's magnitude understanding and, in turn, fraction-related outcomes, supported by prior intervention studies employing similar number line representations.

While the games draw heavily from the science of learning to provide an immersive learning experience, the design ideas have surfaced through a collaborative participatory process of engaging with teachers, students, and researchers to ensure on-the-ground needs and expertise are integrated in our designs, as discussed next.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1.4. Integrating students', teachers', and researchers' expertise through participatory design

Decades of research from laboratory-based science of learning principles offer clear implications for improving educational practice (Richland et al., 2017; Richland, Stigler, & Holyoak, 2012), yet with little to no inroads in schools (Pomerance et al., 2016). A major challenge in applying lab-based research insights into school settings is that they do not account for the everyday challenges, routines, and goals of students and educators (Anderson & Helms, 2001; Brown, 1992; IJzerman et al., 2020; Penuel & Gallagher, 2017).

To address this challenge, we engaged students and teachers in the participatory design process. This process emphasizes user-centered involvement, empowerment, collaborative workshops, and contextual understanding (DiSalvo, Yip, Bonsignore, & Carl, 2017; Simonsen & Robertson, 2013), allowing us to coconstruct educational interventions that are relevant, practical, and tailored to the specific needs and contexts of the school community (Ahn, Campos, Hays, & DiGiacomo, 2019; Sanders & Stappers, 2008). Through this methodology, we open the door to incorporating community voices into research and program development (Anderson-Coto et al., 2024; Bødker, Dindler, Iversen, & Smith, 2022; Gooch et al., 2018). These methodologies encourage close collaboration with practitioners, inspiring context-responsive solutions that address their everyday challenges and build on their existing practices (Cobb, Confrey, DiSessa, Lehrer, & Schauble, 2003; Penuel & Fishman, 2012).

This community-centered and collaborative approach holds the promise of developing more effective strategies that ground mathematical cognition and psychology research in the daily experiences of historically marginalized and underserved communities (Bødker et al., 2022; Galleguillos & Coskun, 2020). Incorporating student voices supports a more comprehensive understanding of school contexts, inspiring pedagogical possibilities that challenge researchers' assumptions and invert research-subject power dynamics (Galleguillos & Coskun, 2020; Warren et al., 2001).

Marrying the science of learning with participatory design principles centers youth and educator voices in the evolution of classroom spaces and educational systems, while utilizing evidence-based learning strategies. By grounding the science of learning research in the experiences and perspectives of students and educators, educational programs can build from the strengths of historically marginalized communities and leverage insights from the science of learning literature (Bermudez et al., 2023). This study offers an example of how scholars can partner with students and educators to combine the science of learning, student voice, and educators' practical expertise through participatory design to create usable and effective educational interventions.

1.5. The current study

This study evaluated the impacts of two games, Fraction Ball and Bottle Caps Bonanza, which were codesigned with students and teachers to promote more advanced student rational number learning, such as fraction addition with unlike denominators. Literature in mathematical cognition and analogical reasoning suggests that advanced rational number concepts are a challenging area to improve. To meet this challenge, we partnered with students and

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

teachers to integrate students' creativity and tacit knowledge of playful learning, teachers' on-the-ground expertise, and researchers knowledge of learning principles from multiple literatures, including analogical reasoning (Richland & Begolli, 2016), the integrated theory of numerical development (Siegler, Thompson, & Schneider, 2011), common misconceptions with fraction concepts (Ni & Zhou, 2005), and playful learning (Hassinger-Das et al., 2018).

Our first aim was to examine the effectiveness of combining Bottle Caps Bonanza and Fraction Ball compared to "business as usual" math instruction. Our second aim was to compare two versions of the games: one condition where equivalent fraction representations were shown simultaneously (e.g., fourths next to eighths) and one condition where the representations were shown sequentially (e.g., first fourths, then eighths). Specifically, the current study addresses the following research questions:

RQ1: Does a combination of Fraction Ball and Bottle Caps Bonanza improve fraction knowledge compared to "business as usual" math instruction?

RQ2: Does the order of presenting representations within the combination of Fraction Ball and Bottle Caps Bonanza impact fraction learning? Specifically, does a simultaneous version (equivalent fraction representations simultaneously visible) improve fraction knowledge compared to a sequential version (equivalent fraction representations shown sequentially)?

RQ3: Are these treatment impacts on students' fraction knowledge moderated by students' prior knowledge and grade level?

Research Questions 1–3 were preregistered when pretest data were partially collected and entered (before all posttest data were entered). To maintain anonymity during the review process, we have included the preregistration as a separate document rather than providing a direct link (see Supplementary Materials, Section C). RQ3 was listed as a secondary ("exploratory") research question in the preregistration. For RQ1, we hypothesized that the intervention would lead to improved fraction knowledge for the treatment group students compared to the comparison group students. We did not preregister hypotheses about RQ2 or RQ3 because, within our team and our teacher partners, there were mixed opinions based on literature and practical experiences with students on which condition would lead to better outcomes. In fact, our team intentionally aimed to choose two conditions for which the direction of the impacts was unclear when compared to each other, but not when compared to business as usual. This was done to ensure that we are testing designs that are poised to benefit students.

2. Method

2.1. Design narrative

In the following section, we present a design narrative illustrating how we partnered with students and educators to iterate on previous versions of our program and extend the activities into the classroom to deepen the learning opportunities and build from the strengths

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and expertise of all design partners (i.e., students, teachers, and researchers). A design narrative is a form of scholarship that describes the origins and process of developing a program or artifact (Bell, Hoadley, & Linn, 2013; Belgrave et al., 2022). These narratives provide meaningful context about the collaboration between different stakeholders in a design process and highlight key aims and design choices. Transparency around this process offers a deeper understanding of the resulting designs and intentionality behind them. It also serves as a model for other scholars, practitioners, and designers to adopt or adapt these practices for their own design projects.

Our codesign process spanned from September 2022 to March 2023, consisting of seven 90-min, on-site codesign sessions with eight teacher partners from three schools in a local Southern California school district. These sessions took place during their afterschool professional development time, with teachers who taught math or special education to fourth- to sixth-grade students. We employed our tiered approach to engagement (Lawrence et al., 2022), starting with level 1 workshopping sessions to generate ideas around extension problems for fourth graders and creating a differentiated version of Fraction Ball for fifth grade. Following each level 1 session, we held smaller, level 2 group sessions for a more in-depth exploration of ideas generated and created a feedback loop with researchers and level 2 teachers that informed and refined lesson activity ideas. Throughout the codesign sessions, we collected video and audio data, observation notes, design artifacts created by teachers, and reflections after the session. In total, we collected 78 min of video recording, 813 min of audio recording, and 15 design artifacts. Findings yielded actionable insights that were then integrated into the intervention.

In parallel to the teacher design sessions, we engaged with a student group to incorporate the youth perspective. The FEMINEERS were an all-female sixth and seventh-grade afterschool engineering club at one of our partner school sites. We challenged the FEMINEERS to design classroom activities that would extend Fraction Ball into the classroom. They had access to a STEM lab space with advanced computing, 3D printing, and a variety of other resources and materials. After a semester of designing and iterating, we invited the FEMINEERS to a maker space in our university campus to present their ideas to a panel of faculty and doctoral students and take a tour of the campus. This 90-min session was audio recorded and transcribed for later analysis.

2.2. Student and teacher contributions through participatory design

Our student design partners, The FEMINEERS, created a low-cost, tabletop game that became central to the classroom implementation and was later named Bottle Caps Bonanza. Although the FEMINEERS had access to state-of-the-art technologies (e.g., 3D printers), the designers of Bottle Caps Bonanza intentionally chose simple, widely available materials: bottle caps, paper, and writing tools (see Table 1 for the study's printed version). Their prototype, inspired by the layout of the Fraction Ball court, emphasized usability and was praised by panelists for its accessibility. As one panelist remarked, "*They wanted to make it as simple as possible… kids can just draw on the floor… and just play. And it's zero cost, which is perfect*." The student presenters also highlighted an extension where teachers could print the

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

boards, "*Once it's graphic[ally] design[ed], you could just send the PDF to teachers and they could print it out*." The students also described gameplay involving two teams flicking bottle caps along the shuffleboard-like game board, and made connections to Fraction Ball on the court:

> *The numbers are like 0/4, 1/4, 2/4, 3/4, and one whole…You flick [the cap] and whatever number you [land on]…you put [the play piece] on the number line [to keep] your score. So if you got zero fourths, like me, you just leave it…And then the other team goes. And this side [of the number line] is…the same thing as a fraction, but in decimals. Like 0.25…it's just the same [as* $\frac{1}{4}$*]…We kind of modeled it off the numbers in the basketball court.*

The prototype embodied key playful learning characteristics and aligned with mathematical cognition research on number lines and magnitude understanding (Bustamante et al., 2020; Link et al., 2013). Its simplicity and adaptability made it well-suited for classroom use and laid a foundation for further iteration with teachers.

Building on the FEMINEERS' design, teacher codesign sessions surfaced refinements that further aligned with science of learning principles, particularly in support of schema building and knowledge transfer. A recurring theme was the need to begin with concrete representations before shifting to abstraction, consistent with the principle of "concreteness fading" (Goldstone & Son, 2005). One teacher emphasized, "*We never give enough time…on the concrete…and we move into the abstract so fast…If we move too quickly into the abstract, they just shut down, and they're really reluctant to ever open back up to fractions again*." Administrators echoed the importance of magnitude understanding and part-whole (or bipartite) relationship.

> *They didn't realize that the numerator… is the counting piece and the denominator is… how it's divided. If they understand that, then it helps when they add and subtract…they can see there are different size pieces that they're dealing with.*

Teachers suggested adapting the court to present multiple denominators side by side, especially "friendly" or proportionally equivalent pairs like 4ths and 8ths, to help visually compare unit sizes. This structure, visible in Table 1, reinforced the principle that with a constant numerator, larger denominators yield smaller pieces.

Teachers also recommended having students create their own courts or boards using familiar denominators:

> *It'd be cool to have [students]…create their own…they can choose fourths or eighths, because that's…easier. You just break everything in half…that will allow you to differentiate and scaffold it for [those] who…need a challenge versus need support.*

One teacher ideated on possible variations inspired by Bottle Caps Bonanza, "You could do [Bottle Caps Bonanza] with 7ths and 9ths and then when you switch sides, maybe you're

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

scoring different amounts… There's all kinds of stuff that kids could extend and create in the class." These recommendations align with analogical reasoning and progressive alignment theories: comparing similar representations (e.g., $\frac{1}{4}$ and 2/8) supports schema development and transfer of knowledge to less familiar contexts (Holyoak, 1983; Kotovsky & Gentner, 1996; Richland & Begolli, 2016).

To further support transfer, teachers proposed embedding repeated opportunities for procedural practice. For example, some proposed having students write number sentences after each play: "*If you're trying to have them look at it and develop a pattern, you might want to have them write [a number sentence] for every single [shot]…they're going to need that repetition and practice*." This practice reflects Rittle-Johnson et al.'s (2015) framework for cyclical integration of conceptual and procedural knowledge, where each reinforces the other through structured, repeated engagement.

Together, these teachers' and students' contributions illustrate how participatory design can produce fraction activities that advance both pedagogical usability and conceptually rich learning. By centering practitioner and youth perspectives, our codesign process offers a promising methodology for integrating science of learning principles into everyday teaching routines, bridging classroom realities with research in ways that support knowledge transfer and address persistent challenges in improving educational outcomes.

The final outputs of codesign with the FEMINEERS and teachers resulted in the Bottle Cap Bonanza game, which offers several important contributions. First, the Bottle Cap Bonanza offered an elegant solution to examine the effects on far transfer for: (a) multiple representations of fractions; (b) simultaneous representations of fractions; and (c) offered an additional play context. All three could theoretically contribute toward enhancing far transfer. Second, the design centers the voices of our student and teacher partners who aimed to create a version of the game that is more equitable by enhancing its accessibility for students who cannot play outdoors due to physical limitations and may not have access to a basketball court. The Bottle Cap Bonanza is highly accessible as it only requires a pen, paper, and a bottle cap. Lastly, teachers asked for classroom activities that reinforce concepts learned outdoors, and the Bottle Cap Bonanza is tightly aligned with the pillars of playful learning (joyful, socially interactive, iterative, actively engaging, and meaningful) and concepts from research in mathematical cognition.

2.3. Intervention components

The intervention consisted of two components: participation in the Fraction Ball court game and the Bottle Caps Bonanza in-class game. In Fraction Ball, students engaged in a basketball game on a redesigned basketball court, rotating through different roles, including shooter, rebounder, and counter. As seen in Table 1, the fraction symbols in each respective area (fourths/eighths and thirds/sixths) reflect possible point values in the game. For example, landing a bottle cap in or making a shot from the $\frac{1}{4}$ area gives players $\frac{1}{4}$ of a point, which is half the distance as 2/4ths of a point. Further, the number lines on the side serve as a scoreboard, which players use to keep track of their points and show both fraction and whole number representations. Player roles reinforce these concepts. For instance, when a shooter successfully

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

made a shot from the 2/4 area, the rebounder announced the corresponding value, "2/4," and the counter moved forward on the number line, adding 2/4 to their team's score. In Fraction Bottle Caps, students took turns using their fingers to flick or shuffle a bottle cap to labeled fraction increments on a tabletop fraction board. See Table 1 for examples of the redesigned courts and the boards.

Both games were called "Make it Count." The objective was for teams of students to earn as many points as possible. Two teams played against each other in the games, and each student had three shots on the court and five in the board game. Teams took turns to shoot or flick, and at the end, they calculated the sum scores and compared their scores.

Two versions of the games were created by providing two designs of the court and board: simultaneous and sequential. In more detail, equivalent fractions (e.g., $\frac{1}{4}$ and 2/8 as well as 2/4 and 4/8) were presented side by side to students in the simultaneous condition, while these fractions were shown sequentially (e.g., fourths first, then eighths) in the sequential condition.

Placing varied rational number notations that share the same magnitudes side-by-side, such as proportional fractions on the scoreboard of the court and board (e.g., $\frac{1}{4}$ and 2/8), has additional benefits. For example, children may erroneously posit that 1/4 is smaller than 1/10, likely because they consider the magnitude of the whole number in the denominator (Fazio et al., 2016; Mazzocco & Devlin, 2008). By placing equivalent fractions, for example, $\frac{1}{4}$ and 2/8ths, next to each other on the board, reasoners are prompted to engage in a process of contrasting the different numbers and draw relationships to the equivalent magnitude they represent on the court/board and scoreboard (number line), which also shows that, for example, 3/8ths is smaller than 2/4ths.

The current intervention comprised two Bottle Caps Bonanza sessions and two Fraction Ball sessions. In the first Bottle Caps Bonanza session, students either used simultaneously visible or sequentially visible boards with fourths and eighths. In the second session, students played the same games on the third and sixth boards. Similarly, in Fraction Ball, students played on the fourths and eighths side of the court in the first session and on the thirds and sixths side in the second session (see more detail in Table 1 and the Supplementary Materials, Section B).

2.4. Participants

Participants included 10 teachers and a total of 241 fourth ($N = 201$) and fifth ($N = 40$) graders from the Santa Ana Unified School District in California. The 241 students were predominantly Hispanic or Latino ($n = 232$), with 49% female students ($n = 118$) and 51% male students ($n = 123$). Forty-eight percent ($n = 116$) of the students were identified as English Language Learners. Over 70% of the students received free or reduced lunch. Thirteen percent received special education services. There were seven treatment classrooms: 87 students were in the simultaneous condition, and 79 students were in the sequential condition. There were three additional comparison classrooms with 75 students in total. Among fourth graders, 65 were assigned to the simultaneous condition, 63 to the sequential condition. Seventy-three were in the comparison group. Fifth graders were similarly distributed, with 14 students in

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

each treatment condition and 12 in the comparison group. Overall, group sizes were relatively balanced within each grade level.

This study, including the recruitment procedure, was approved by the University of California, Irvine institutional review board. To be included in the study, participants were required to be math teachers in either fourth or fifth grade within the Santa Ana Unified School District. Teachers were invited to participate through email and were informed that an experiment would be conducted to assess the efficacy of the math games. Then, teachers who signed up were informed that their students within their classroom would be randomly assigned to either the simultaneous or sequential condition. Three additional teachers enrolled as comparison group teachers, and their students only took part in the pretest and posttest. Teachers were provided incentives for their involvement ($50 per session). Parents of students in all participating classes (both treatment and comparison groups) received information letters about the intervention from the teachers.

2.5. Study design

The experiment primarily employed an A/B test design, where students in the seven treatment classrooms were randomly assigned within their classrooms to either the simultaneous or sequential condition. Four math periods were dedicated to the intervention: two sessions were Fraction Ball court activities, and the other two were Bottle Caps Bonanza in-class activities. There were no consecutive court or in-class activities. In other words, either the first and third sessions or the second and fourth sessions were in-class sessions (same for court sessions).

We balanced the class-court activity order. Specifically, among the seven sequential groups, three groups ($N_{\text{Students}} = 42$) started with court Fraction Ball activities, and their sequence followed a court-class-court-class pattern. The remaining four groups ($N_{\text{Students}} = 45$) initiated the intervention with in-class Bottle Caps Bonanza activities, following a class-court-class-court sequence. Similarly, among the seven simultaneous groups, four groups ($N_{\text{Students}} = 41$) started with court Fraction Ball activities, while the other three groups ($N_{\text{Students}} = 38$) initiated with in-class Bottle Caps Bonanza activities. In addition, the three comparison classes ($N_{\text{Students}} = 75$) only took the pretest and posttest of fraction knowledge and maintained their regular math classes during the intervention.

2.6. Procedure

A one-class period pretest was administered within 2 weeks prior to the start of the intervention. The research team conducted all the intervention sessions. The four intervention sessions were delivered over a period of 7−14 calendar days, with an average of approximately 11 days. The interval between sessions ranged from 1 to 7 days, with an average interval of approximately 4 days between sessions. Lead principal investigators conducted training sessions for graduate students and research assistants before the intervention. Researchers (principal investigators, graduate students, and full-time research staff) with experience in teaching and/or leading Fraction Ball activities led the intervention sessions. Research assistants provided additional support during the sessions. For in-class activities, we ensured the

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

presence of at least one researcher and one research assistant. During court activities, there were a minimum of two researchers and two research assistants present on the court. While teachers were encouraged to assist with either court or in-class activities, it was not mandatory for them to deliver the intervention. The intervention fidelity data were not collected because the sessions were fully scripted (see the lesson scripts in the Supplementary Materials, Section B) and delivered by the research team. Finally, students in the treatment group completed the posttest within 1−2 days of their final intervention session. Specifically, six of the seven treatment classrooms administered the posttest 1 day after the final session, and one classroom administered it 2 days later. For the comparison group, three classrooms conducted the posttest on the same two dates most commonly scheduled for the treatment classrooms.

2.7. Measures

2.7.1. Fraction knowledge test

Identical preintervention and postintervention tests of fraction knowledge with both English and Spanish instructions were administered to all students. We created two versions of the tests by counterbalancing the timed items. Specifically, we alternated the order of the first (i.e., items 1–12) and the second half (i.e., items 13–24) of the timed items. One version was administered to half of the classrooms ($N_{\text{Class}} = 5$) and the other version to the other half of the classrooms ($N_{\text{Class}} = 5$). Classes received the same version for pretest and posttest. However, since the test version was not randomized at the student level, the class could be confounded with the test version when comparing the treatment versus comparison, because the treatment was assigned at the classroom level, but not when comparing the two treatment conditions to each other, because these were randomly assigned within classrooms. To address this potential confound, we included the test version as an additional covariate, along with the pretest average composite and grade, when estimating the treatment versus comparison effects as a robustness check, which was not preregistered.

The test assessed fraction addition, magnitude comparison, and number line estimation. There were 46 items in total, with 24 timed items and 22 untimed items. Timed subtests were included to assess fluency and efficient processing of fractions, while untimed subtests were designed to evaluate children's conceptual understanding and procedural knowledge (Begolli et al., 2024). Including both allows for a more comprehensive assessment of students' fraction knowledge. Part of the fraction addition and number line estimation items were selected from the recent Fraction Ball experiment (Guo et al., 2024). The fraction magnitude comparison and number line with hatch mark estimation items were added to better capture the outcomes of the current version of the intervention. The current test consisted of nine mutually exclusive subtests (three timed subtests and six untimed subtests) that represent different aspects of fraction knowledge.

2.7.1.1. Timed subtests: In the first portion of the test, students were asked to finish as many items as possible out of the total 24 items within 3 min. These 24 items consisted of three subtests: *timed fraction addition with equal denominators* (8 items), *timed fraction addition with unequal denominators* (8 items), and *timed fraction*

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

magnitude comparison (8 items). The timed items were not grouped by their subtest type in the test. They were presented in an interleaved sequence. The internal consistency for the timed items is 0.88 at pretest and 0.90 at posttest.

2.7.1.2. Untimed subtests: The second section was an untimed test, and students were informed that they had as much time as they wanted in the remaining class period (approximately 45 min) to complete the test. All students completed the untimed subtest within the allotted time. The untimed portion included six subtests: *untimed blank 0−1 number line estimation* (3 items), *untimed blank 0−5 number line estimation* (3 items), *untimed 0−1 number line with hatch marks estimation* (3 items), *untimed 0−5 number line with hatch marks estimation* (3 items), *untimed fraction addition with unequal denominators* (6 items), and *untimed fraction magnitude comparison* (4 items). Items in these six subtests were given in a fixed order and grouped according to their respective subtests. The internal consistency for the untimed items is 0.90 at both pretest and posttest.

2.7.1.3. Near and far transfer items: Additionally, we categorized the test items into near transfer (26 items) and far transfer (20 items). Near transfer items consisted of numbers and procedures that students were directly exposed to during the four intervention sessions. More specifically, these items included fractions that were presented in thirds, fourths, sixths, and eighths. In contrast, far transfer items involved numbers and procedures that were not directly presented during the intervention. We have included halves, fifths, ninths, tens, and twelfths in far transfer items. The Section A, Table A1 in the Supplementary Materials, shows the pretest and posttest items by subtests and type of transfer.

2.7.2. Item scoring

Number line estimation items (both blank number line and number line with hatch marks items) were scored using percent absolute error (PAE; Booth & Siegler, 2006). PAE was operationalized as [absolute value of the length of (student's response – correct response)]/ (length of the entire number line). We then applied a natural log transformation to the PAE values and reverse-coded to allow larger values to indicate more accurate responses. All other subtest items were dichotomously scored as correct (assigned a value of 1) or incorrect (assigned a value of 0). Fraction answers were considered correct as long as the values were accurate, regardless of their specific forms.

For items that were skipped and left blank, we adopted the common practice among teachers and scored them as incorrect (assigned a value of 0). An exception was made for the number line estimation items (including all four subtests: blank 0−1, blank 0−5, 0−1 with hatch marks, and 0−5 with hatch marks). In cases where students left the number line items blank, we imputed the 90th percentile PAE value specific to that item. In other words, the imputed 90th percentile PAEs were used to "represent" the missing values and were tailored to students' performance on the respective number line estimation item. A value of 0 (indicating incorrectness for other items) was not imputed, as it would indicate a perfect correct response. Conversely, a large PAE (e.g., at the 90th percentile) would imply poor performance, consistent with our approach to missingness for other items.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.7.3. Composites

First, we computed average raw scores for each of the nine subtests at both pretest and posttest by calculating the means of the items corresponding to each subtest. Then, the average raw score for each subtest was standardized using the average grade standard deviation from both Grade 4 and Grade 5. For pretest scores, the standard deviations were from the full sample, while for posttest scores, only those from the comparison group were used. Finally, average pretest and posttest overall composites were calculated by taking the means of the standardized scores across the nine subtests. Descriptive statistics of pretest and posttest composite scores can be found in the Supplementary Materials, Section A, Table A2.

For each subtest, we also calculated the average raw scores for both near transfer items and far transfer items. Then, we standardized near and far transfer scores using the same procedure as described above. The overall near and far transfer composites were created by averaging the standardized scores across all the subtests.

2.8. Data analysis plan

2.8.1. Baseline equivalence

First, we assessed the equivalence between the treatment and comparison groups as well as between the simultaneous and sequential groups on measured variables. To investigate the equivalence between the treatment and comparison groups, we regressed students' grade level, attrition from pretest to posttest, attending pretest and posttest, and raw subtest scores at pretest and posttest on treatment status (treatment or comparison; with comparison as the reference group) in separate models, clustering standard errors at the teacher level. For the equivalence between the simultaneous and sequential groups, we restricted our sample to the treatment group students and regressed the aforementioned variables on treatment conditions (simultaneous or sequential; with sequential as the reference group).

2.8.2. Overall treatment effects

We estimated the treatment (both conditions) versus comparison effects following the preregistered model specifications (RQ1). In our preferred models, we employed regression models regressing the outcome variable (Posttest) on the treatment status (Treatment), while controlling for pretest average composite (PreAverage) and grade level (Grade 4), with standard errors clustered at the teacher level to account for the nonindependence of students within classrooms. Our regression equation is as follows:

$$\text{Posttest} = \beta 0 + \beta 1\ \text{Treatment} + \beta 2\text{PreAverage} + \beta 3\ \text{Grade4} + \varepsilon\ (\text{standard errors were clustered at the teacher level})$$

$\beta 1$ represents the estimated effect of the intervention on students' fraction knowledge, holding their grade level and pretest average composite constant.

As robustness checks, we preregistered to first utilize generalized linear mixed models (GLMMs) with random intercepts by teachers. In our preferred models with clustered standard errors at the teacher level, the intercepts were calculated as the average standard error across all classrooms, assuming there is no individual classroom-level difference in terms of

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

standard errors. Nevertheless, within GLMMs, intercepts are determined by calculating the average standard error along with a measure of variability specific to each classroom. We also preregistered to employ regression models using full information maximum likelihood (FIML) to address missing data and estimate treatment effects for the entire sample.

Moreover, we also added two rounds of robustness checks that we did not preregister. First, as shown in our preregistered preferred model results (Table 3), the standard errors associated with a few subtest outcomes were very large (e.g., timed fraction addition with equal and unequal denominators, untimed fraction addition with unequal denominators, and estimation on both 0–5 blank number line and number line with hatch marks). One potential explanation for these large standard errors could be that the pretest average composite, which we utilized as a covariate in the preferred model, may not be strongly correlated with these specific subtest scores. To mitigate this, we included individual subtests into the model, in addition to the pretest average composite and grade level. In more detail, for each posttest subtest outcome, the corresponding subtest measured at pretest was added as an additional covariate, along with the pretest average composite and grade level. Second, as mentioned in the *Measures* subsection, we introduced the test version as an additional covariate, while also controlling for the pretest average composite and grade level, to further assess the robustness of our estimates.

2.8.3. Simultaneous versus sequential effects

In investigating the simultaneous versus sequential effects (RQ2), we preregistered to exclusively include the treatment group students. The analysis involved regressing the outcome variable (Posttest) on the assignment to the simultaneous condition (Simultaneous; with the sequential group as the reference group) while controlling for pretest average composite (PreAverage) and grade level (Grade 4). We did not cluster standard errors as the students were randomized within classrooms to either the simultaneous or sequential condition. The regression equation is as below:

$$\text{Posttest} = \beta 0 + \beta 1\ \text{Simultaneous} + \beta 2\ \text{PreAverage} + \beta 3\ \text{Grade4} + \varepsilon$$

$\beta 1$ is the estimated effect of the simultaneous condition on students' fraction knowledge compared to the sequential condition, holding their grade level and pretest average composite constant.

As robustness checks for missingness, as preregistered, we utilized FIML to estimate the results based on the full treatment group sample. Additionally, we conducted two rounds of robustness checks that were not preregistered. First, as shown in Table 4, the standard errors for most subtest estimates were large. Thus, similar to the robustness check for estimating the overall treatment effects, we controlled for individual pretest subtest scores in addition to the pretest average composite and grade level, to assess if it could lead to a reduction in these standard errors. Second, when comparing the pretest scores between the two treatment groups, we observed that the pretest 0–5 blank number line estimation subtest was already statistically significant (Table 2). Therefore, we also added this pretest subtest score as a covariate, along with the pretest average composite and grade level, to test the sensitivity of our estimates.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
Descriptive statistics of student demographics and raw scores at pretest and posttest

	Full sample			Comparison			Treatment-simultaneous			Treatment-sequential						
Variable	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	b_1	p_1	b_2	p_2
Fourth grade	241	83%		75	84%		79	82%		87	84%		−0.01	.97	−0.02	.78
Fifth grade	241	17%		75	16%		79	18%		87	16%		0.01	.97	0.02	.78
Attrition from pretest to posttest	229	1.0%		71	3.0%		75	0%		83	0%		−0.03	.05*	0	/
Attending pretest	241	95%		75	95%		79	95%		87	95%		.001	.92	−0.00	.89
Attending posttest	241	99%		75	97%		79	100%		87	100%		0.03	.04*	0	/
Pretest																
Timed fraction addition with equal denominators	229	58%	30%	71	64%	30%	75	55%	32%	83	56%	27%	−7.86	.15	−1.64	.73
Timed fraction addition with unequal denominators	229	14%	29%	71	27%	37%	75	6.3%	19%	83	8.4%	26%	−19.68	.27	−2.10	.56
Timed fraction magnitude comparison	229	43%	27%	71	55%	28%	75	38%	25%	83	38%	26%	−16.80	.09	0.06	.99
Untimed fraction addition with unequal denominators	228	21%	37%	71	37%	46%	74	15%	31%	83	12%	29%	−23.50	.35	2.84	.55
Untimed fraction magnitude comparison	228	57%	37%	71	67%	39%	74	56%	32%	83	51%	37%	−12.69	.41	4.88	.39
PAE blank 0–1 number line	229	17%	18%	71	14%	15%	75	18%	16%	83	19%	21%	4.51	.37	−1.28	.67
PAE blank 0–5 number line	229	21%	13%	71	20%	14%	75	23%	14%	83	19%	11%	1.41	.78	4.22	.03*
PAE 0–1 number line with hatch marks	229	17%	14%	71	11%	13%	75	20%	15%	83	19%	13%	8.08	.15	0.43	.85
PAE 0–5 number line with hatch marks	228	14%	15%	71	12%	15%	74	15%	14%	83	14%	15%	2.26	.64	1.33	.56
Posttest																
Timed fraction addition with equal denominators	239	67%	31%	73	69%	31%	79	64%	33%	87	70%	28%	−1.72	.85	−5.46	.25
Timed fraction addition with unequal denominators	239	21%	34%	73	35%	41%	79	15%	30%	87	13%	28%	−21.21	.32	1.83	.69
Timed fraction magnitude comparison	239	53%	27%	73	60.%	30%	79	49%	27%	87	51%	25%	−10.44	.38	−2.72	.50
Untimed fraction addition with unequal denominators	239	27%	40%	73	42%	44%	79	24%	37%	87	18%	34%	−21.13	.34	6.04	.28
Untimed fraction magnitude comparison	239	61%	37%	73	70%	38%	79	57%	35%	87	56%	36%	−13.39	.30	0.32	.95
PAE blank 0−1 number line	239	15%	17%	73	14%	17%	79	16%	17%	87	14%	16%	1.64	.73	1.94	.46
PAE blank 0−5 number line	239	21%	14%	73	21%	13%	79	22%	14%	87	20%	14%	0.25	.96	1.57	.47
PAE 0–1 number line with hatch marks	239	14%	13%	73	12%	13%	79	16%	13%	87	14%	12%	3.49	.48	1.10	.58
PAE 0–5 number line with hatch marks	239	12%	13%	73	13%	15%	79	12%	14%	87	10%	10%	−2.25	.66	2.88	.13

Note. b_1 and p_1 are based on regressing each variable on either the treatment (simultaneous or sequential) or comparison group for the full sample, clustering standard errors by teacher. b_2 and p_2 are based on regressing each variable on either the simultaneous or sequential group for the treatment subsample. The *N* in the Variable section refers to the total sample possible, including student attrition.

*$p < .05$.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.8.4. Moderation of treatment effects

Following our preregistration, we tested if students' prior knowledge level (as measured by the pretest of fraction knowledge) moderated the treatment effects (RQ4). In particular, we were interested in exploring if students with varying prior fraction knowledge levels would benefit differently from different treatment conditions. The moderation of prior knowledge was assessed using both a dichotomous variable indicating high prior knowledge and a continuous variable of prior knowledge (pretest average composite). To create the dichotomous variable for high prior knowledge, we performed a median split across all students ($n_{\text{high}} = 115$; $n_{\text{low}} = 114$) based on their pretest average composite when estimating its moderating role on the treatment versus comparison effects. When examining the moderating role of prior knowledge on the simultaneous versus sequential effects, the median split was performed within the treatment group students ($n_{\text{high}} = 79$; $n_{\text{low}} = 79$). To better present the potential differential effects, we also reported the estimated treatment impacts for the high and low prior knowledge subgroups. In addition, we explored the potential effects of student grade level (Grade 4 vs. Grade 5).

Specifically, we included an interaction term between the group assignment (i.e., treatment vs. comparison; simultaneous vs. separate) and the potential moderators (i.e., prior knowledge level and grade level) in the regression models. When testing the moderating effects on the treatment versus comparison impact on the posttest overall composite, the regression equation is:

$$\begin{aligned}\text{Posttest} = {} & \beta 0 + \beta 1\ \text{Treatment} + \beta 2\ \text{PreAverage} + \beta 3\ \text{Treatment} \\ & * \text{Moderator}\left(\text{prior knowledge level or grade level}\right) + \beta 4\ \text{Grade4} \\ & + \varepsilon\ (\text{standarderrorsclusteredattheteacherlevel})\end{aligned}$$

Finally, we examined the moderating effects of the simultaneous versus sequential condition impact, and followed the regression equation below:

$$\begin{aligned}\text{Posttest} = {} & \beta 0 + \beta 1\ \text{Treatment} + \beta 2\ \text{PreAverage} + \beta 3\ \text{Simultaneous} \\ & * \text{Moderator}\left(\text{priorknowledgelevelorgradelevel}\right) + \beta 4\ \text{Grade4} + \varepsilon\end{aligned}$$

3. Results

3.1. Experiment results

Table 2 presents the descriptive statistics of student demographics as well as pretest and posttest raw scores by subtests. There were 12 missing pretests, evenly distributed across three groups (four in the simultaneous group, four in the sequential group, and four in the comparison group). Our absolute attrition rate from pretest to posttest was low, less than 1% (two out of 229 students). As both of the students who were missing the posttest were from the comparison group, we found the treatment group exhibited statistically significantly lower attrition from the pretest to the posttest ($b = -0.03$, $p < .05$), and more students attending the

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3
Estimated treatment (both conditions) versus comparison effects

	No covariates ($N = 239$)			Pretest average composite and grade covariates ($N = 227$)		
Standardized posttest outcome	*b*	*(SE)*	*p*	*b*	*(SE)*	*p*
Overall composite	−0.26	(0.39)	.51	0.15	(0.04)	.01**
Near transfer composite	−0.27	(0.37)	.48	0.12	(0.06)	.08
Far transfer composite	−0.29	(0.38)	.46	0.12	(0.04)	.01**
Timed fraction addition with equal denominators	−0.05	(0.24)	.85	0.26	(0.14)	.10
Timed fraction addition with unequal denominators	−0.53	(0.50)	.32	−0.14	(0.16)	.39
Timed fraction magnitude comparison	−0.37	(0.41)	.38	0.06	(0.05)	.23
Untimed fraction addition with unequal denominators	−0.52	(0.52)	.34	−0.09	(0.16)	.59
Untimed fraction magnitude comparison	−0.34	(0.31)	.30	0.10	(0.08)	.28
Reversed natural log of PAE blank 0−1	−0.10	(0.32)	.76	0.36	(0.05)	.00***
Reversed natural log of PAE blank 0−5	0.01	(0.39)	.97	0.39	(0.12)	.01*
Reversed natural log of PAE 0−1 with unequal denominator hatch marks	−0.41	(0.48)	.41	0.08	(0.07)	.33
Reversed natural log of PAE 0−5 with unequal denominator hatch marks	−0.07	(0.44)	.88	0.37	(0.14)	.03*

Note. Standardized scores are reported to allow for comparison across measures. The outcome variables were only regressed on treatment status (comparison group as the reference group) in the first model. The second model is our preregistered preferred model, which includes pretest average composite and grade as covariates. The pretest average composite is the average z-score of student performance on the pretests using the average grade standard deviation from Grade 4 and Grade 5 for all participants. The posttest overall composite was calculated using the average grade standard deviation for the comparison group. Percent absolute errors (PAEs) on the number lines were transformed using natural log and reverse-coded so that positive scores indicate better performance. Clustered standard errors by teachers are in parentheses.

*$p < .05$, **$p < .01$, ***$p < .001$.

posttest ($b = 0.03$, $p = .04$). No other demographics were statistically significantly different between the treatment and comparison groups nor between the simultaneous and sequential groups. In terms of pretest raw scores, students in the simultaneous group demonstrated statistically significantly higher PAEs than those in the sequential group for the 0–5 blank number line estimation subtest ($b = 4.22$, $p = .03$).

3.1.1. Estimated treatment versus comparison effects (RQ1)

3.1.1.1. Preferred model: We first estimated the treatment (both conditions) versus comparison effects on students' fraction knowledge. In Table 3, the first model provides results without any covariates, and we present these estimates to explore the sensitivity of the treatment coefficients when covariates are introduced. The second model represents our preregistered preferred model, where we included average pretest composite and student grade level as covariates.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

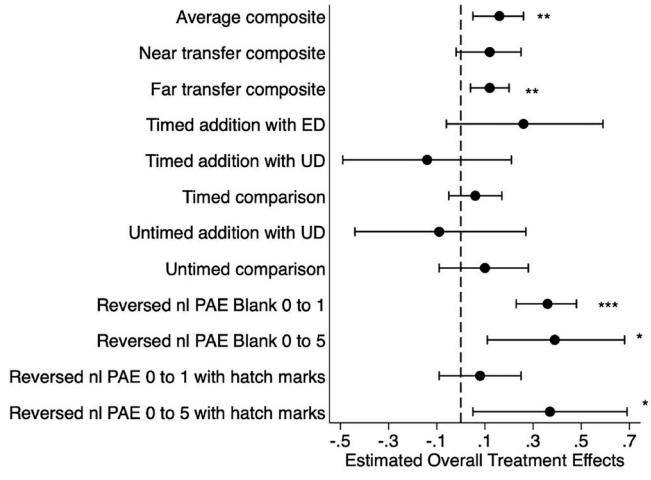


Fig. 1. Treatment versus control estimated effects by standardized outcome.
Note. $N = 227$. ED, equal denominators; nl, natural log; UD, unequal denominators. Estimates represent regression coefficients for treatment status. The estimates can be interpreted as standardized treatment effects in pooled grade SDs of the comparison group. Models are specified using clustered standard errors by teachers, controlling for pretest average composite and grade level. Bars show 95% confidence intervals. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 3 indicates that, before adjusting for pretest scores, estimates are extremely imprecise, with standard errors generally exceeding 0.30. After controlling for pretest average composite and grade, estimates were far more precise, with standard errors falling to mostly below 0.10. We thus focus our discussion on the adjusted models, which were also our preregistered specification. Statistically significant impacts on the overall composite were observed when comparing the treatment groups (both treatments combined) to the comparison group ($b = 0.15$, $p < .01$). Additionally, we reported the estimated effects on the near/far transfer composite and subtests to provide a comprehensive understanding of how improvements vary across different aspects of fraction knowledge. We observed a positive impact on the far transfer composite ($b = 0.12$, $p < .01$). Moreover, the largest gains were on the number line estimation subtests, including 0–1 blank number line estimation ($b = 0.36$, $p < .001$), 0–5 blank number line estimation ($b = 0.39$, $p < .01$), and 0–5 number line with hatch marks estimation ($b = 0.37$, $p = .03$). Fig. 1 illustrates these estimated treatment impacts on the various composites and subtests. In addition, students' raw test scores' changes from pretest to posttest by individual classrooms and treatment status also suggests that students in the treatment groups showed steeper growth in their fraction knowledge, compared to the students in the comparison group (as shown in the Supplementary Materials, Section A, Fig. A1).

3.1.1.2. Robustness checks: First, we present the estimated treatment versus comparison effects based on GLMMs in the Supplementary Materials, Section A, Table A3. These estimates exhibited similar magnitudes to those from our preregistered preferred model, which includes pretest average composite and grade as covariates and clustered standard errors by teacher. Nevertheless, the pattern of statistical significance/nonsignificance changed for two

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

outcomes: the near transfer composite ($b = 0.12$, $p < .01$) and timed addition with equal denominator ($b = 0.26$, $p = .02$) became statistically significant. Section A, Table A4 in the Supplementary Materials, shows estimated fixed and random effects components of the GLMMs. We present three models: the baseline model without any predictors, the model with only treatment status as a predictor (no other covariate), and our preferred model with the pretest average composite and grade covariates. The baseline model results indicated a statistically significant variability in students' posttest average composite scores at both the teacher and student levels. In addition, it revealed that 30% of the residual variance in the overall posttest composite was due to differences between teachers. After accounting for the pretest average composite and grade, the residual variance attributed to teacher differences decreased to zero, indicating that conditional on intervention status, children in the same classroom did not systematically change together from pretest to posttest. The variability of students within classrooms remained statistically significant, but was significantly reduced in magnitude.

Moreover, we present the results of additional robustness checks in the Supplementary Materials, Section A, Table A5 alongside the outcomes of our preregistered preferred model (the first model) to facilitate a more comprehensive comparison. The estimated impacts from the models employing FIML are outlined in the second model. Most estimates closely align with the results based on our preregistered preferred model. A shift in the pattern of significance was observed for the near transfer composite: the estimated impact on the near transfer composite ($b = 0.12$, $p = .045$) became statistically significant in this model. In addition, as shown in the third model, when accounting for the individual pretest subtest scores in addition to the pretest average composite and grade, most of the standard errors only showed marginal reductions. The estimated impact on the 0–5 number line with hatch marks estimation subtest ($b = 0.19$, $p = .03$) became statistically significant. In the fourth model, we incorporate the test version as an additional covariate. In comparison to our preferred model, the estimates remained highly similar, with a slight decrease in some of the standard errors. The estimated effect on the timed addition with equal denominators subtest ($b = 0.21$, $p = .01$) became statistically significant.

3.1.2. Estimated simultaneous versus sequential effects (RQ2)

3.1.2.1. Preferred model: In Table 4, we report the estimated simultaneous versus sequential effects on students' fraction knowledge following our preregistration. We report the results from both the model without covariates (the first model) and our preregistered preferred model with average pretest composite and grade covariates (the second model). As shown in the table, the overall posttest composite was not statistically significantly different between the simultaneous and sequential groups. The simultaneous group outperformed the sequential group on untimed fraction addition with unequal denominators, though this difference was not statistically significant ($b = 0.21$, $p = .05$). Fig. 2 presents these estimated effects on the various aspects of fraction knowledge. Given the number of comparisons and the benefits not reaching conventional thresholds for statistical significance of this estimate, we cannot generalize beyond our sample and recommend interpreting them with caution.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 4
Estimated simultaneous versus sequential condition effects

	No covariates ($N = 166$)			Pretest average composite and grade covariates ($N = 158$)		
Standardized posttest outcome	*b*	*(SE)*	*p*	*b*	*(SE)*	*p*
Overall composite	−0.03	(0.10)	.72	0.02	(0.05)	.66
Near transfer composite	−0.03	(0.09)	.73	0.02	(0.05)	.71
Far transfer composite	−0.03	(0.10)	.75	0.03	(0.05)	.63
Timed fraction addition with equal denominators	−0.15	(0.13)	.25	−0.12	(0.12)	.30
Timed fraction addition with unequal denominators	0.05	(0.11)	.69	0.07	(0.08)	.41
Timed fraction magnitude comparison	−0.10	(0.14)	.50	−0.04	(0.11)	.73
Untimed fraction addition with unequal denominators	0.15	(0.14)	.28	0.21	(0.11)	.05
Untimed fraction magnitude comparison	0.01	(0.14)	.95	0.09	(0.11)	.44
Reversed natural log of PAE blank 0−1	−0.10	(0.15)	.51	−0.02	(0.12)	.84
Reversed natural log of PAE blank 0−5	0.01	(0.16)	.97	0.02	(0.15)	.87
Reversed natural log of PAE 0−1 with unequal denominator hatch marks	−0.07	(0.16)	.68	0.02	(0.13)	.89
Reversed natural log of PAE 0−5 with unequal denominator hatch marks	−0.11	(0.14)	.43	−0.03	(0.11)	.78

Note. Standardized scores are reported to allow for comparison across measures. Standard errors are in parentheses. Both models were restricted to treatment group students. The outcome variables were only regressed on treatment group assignment (sequential group as the reference group). The second model is our preregistered preferred model, which includes pretest average composite and grade covariates. The pretest average composite is the average z-score of student performance on the pretests using the average grade standard deviation from Grade 4 and Grade 5 for all participants. The posttest overall composite was calculated using the average grade standard deviation for the comparison group. Percent absolute errors (PAEs) on the number lines were transformed using the natural log and reverse-coded so that positive scores indicate better performance.

3.1.2.2. Robustness checks: We present the outcomes of robustness checks along with those from our preregistered preferred model (the first model) in the Supplementary Materials, Section A, Table A6. We included three additional models as robustness checks: the model using FIML and including the full sample, the model adding pretest individual subtest scores at pretest, and the model controlling for pretest subtest scores of 0–5 blank number line estimation. None of the estimates changed significantly compared to those from our preferred model when evaluating the simultaneous versus sequential effects.

3.1.3. Moderation of treatment effects (RQ3)

We explored if students' prior knowledge and grade level moderated the treatment versus the comparison effects as well as the simultaneous versus sequential effects. We found that prior knowledge moderated the simultaneous versus sequential effects. As shown in Table 5, after creating a dichotomous variable for high prior knowledge via median split based on student pretest average composite, we found the interaction term between the prior knowledge

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

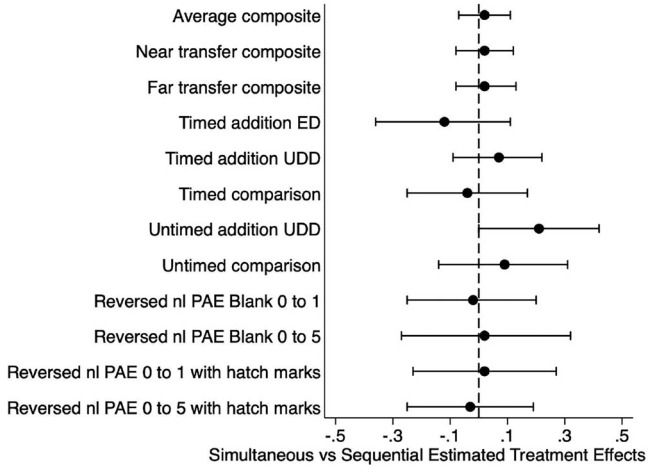


Fig. 2. Estimated simultaneous versus sequential condition treatment effects.
Note. $N = 158$. ED, equal denominators; nl, natural log; UD, unequal denominators. Estimates represent regression coefficients for simultaneous versus sequential condition status with the sequential condition as the reference group, controlling for pretest average composite and grade level. The estimates can be interpreted as standardized treatment effects in pooled grade SDs of the comparison group. Bars show 95% confidence intervals.

Table 5
Estimated moderation effects of student prior knowledge on the simultaneous versus sequential effects

Model	*b*	*(SE)*	*p*
High and low prior knowledge group			
High prior knowledge	−0.04	(0.09)	.63
Simultaneous group	−0.12	(0.07)	.08
High prior knowledge X Simultaneous group	0.27	(0.09)	.00**
Fifth grade	0.04	(0.07)	.52
Pretest average composite	0.86	(0.07)	.00***
Prior knowledge as continuous variable			
Pretest average composite	0.81	(0.06)	.00***
Simultaneous group	0.05	(0.05)	.28
Pretest average composite X Simultaneous group	0.23	(0.08)	.01**
Fifth grade	0.04	(0.07)	.58

Note. Standardized scores are reported to allow for comparison across measures. Both models were used to estimate the moderation effects of student prior knowledge (as measured by pretest average composite) on the simultaneous versus sequential effects. In the first model, a dichotomous variable of prior knowledge was used in the interaction term. In the second model, the pretest average composite as a continuous variable of prior knowledge was used in the interaction term. The pretest average composite is the average z-score of student performance on the pretests using the average grade standard deviation from Grade 4 and Grade 5 for all participants. The posttest overall composite was calculated using the average grade standard deviation for the comparison group.
$^{**}p < .01$, $^{***}p < .001$.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

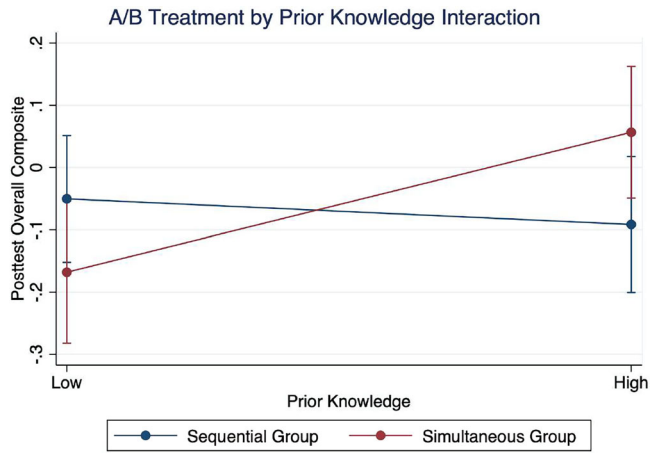


Fig. 3. Student prior knowledge moderating simultaneous versus sequential condition treatment effects.
Note. The two prior knowledge groups ($n_{high} = 79$; $n_{low} = 79$) were created using a median split of high and low groups using pretest average composite.

level and simultaneous group assignment was statistically significant ($b = 0.27$, $p < .01$). More specifically, as demonstrated in Fig. 3, students with higher prior knowledge benefited more from the simultaneous condition and the simultaneous condition had a statistically significant positive impact on students with above median pretest scores (*Simultaneous* $b = 0.15$, $p = .03$), but had a nonsignificant negatively signed impact on students with below median pretest scores (*simultaneous* $b = -0.12$, $p = .08$). Moreover, this finding holds true when treating prior knowledge as a continuous variable, directly employing pretest average composite in the interaction term (*Pretest Average Composite X Simultaneous Group* $b = 0.23$, $p < .01$).

However, we did not find student prior knowledge moderated the treatment versus comparison effects ($p > .05$; Supplementary Materials, Section A, Table A7), nor did student grade level moderate treatment versus comparison or simultaneous versus sequential impacts (both ps $> .05$; Supplementary Materials, Section A, Table A8).

4. Discussion

The results of our study suggest that the iterative redesign of the original Fraction Ball intervention was successful. We will describe the main findings here and discuss the results in relation to the research questions in more detail. Students in the two play-based conditions combined (simultaneous and sequential orders) demonstrated greater improvements in their overall rational number knowledge ($b = 0.12$) and far transfer ($b = 0.15$) compared to students who continued with business-as-usual math class. An examination of the constructs within the posttest measure indicated that these impacts may have been driven by improvements on estimation tasks on the blank number lines and number lines with hatch marks,

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

which yielded moderate effects (effect sizes 0.36−0.39). When comparing across play conditions, students in the simultaneous group showed higher performance compared to those in the sequential condition on untimed addition tasks involving unlike denominators, though this difference was not statistically significant ($b = 0.21$, $p = .05$). However, no discernible differences were observed across other outcome measures. We also found that students with higher prior knowledge demonstrated greater benefits from simultaneous presentations compared to sequential presentations ($b = 0.15$, $p = .03$). Yet, prior knowledge did not moderate the impacts of being in either treatment condition in comparison to business-as-usual math class. We discuss the results next in relation to the research questions.

4.1. Design process

We first speak to the design of the intervention itself. By engaging students and teachers in codesign, we were able to develop an intervention that builds on teacher intuitions and science of learning principles to bring decades of research into the educational arena in a novel and usable format. Teachers' and students' insights and design recommendations were highly aligned with several science of learning principles of comparison, including analogical reasoning, "concreteness fading," and the integrated theory of numerical development.

This was reflected by the new game design presented by the FEMINEER students, who maintained core science of learning principles present in previous iterations of Fraction Ball (e.g., number line representations embedded on court and scoreboard). In addition, teachers argued that students need more practice with concrete representations before moving on to abstract formulas, invoking the principle of "concreteness fading" (Goldstone & Son, 2005).

Teachers also stressed the importance of including more attention to the part/whole relationship between the numerator and denominator in our game designs. In this sense, they were challenging the research team to develop a design that brings together two distinct theoretical accounts of rational number development. Namely, understanding fractions as magnitudes represented on a number line (Siegler, Thompson & Schneider, 2011) and on the other hand, conceptualizing fractions as bipartite relations between the numerator and denominator. First, Siegler and colleagues' theory suggests that physical number lines are an important tool for building fraction magnitude understanding and in turn reduce common misconceptions (e.g., whole number bias). In contrast, Dewolf et al. (2015) posit that fractions are best understood as inherently two-dimensional due to their bipartite discrete (i.e., countable) numerator and denominator elements, which they contrast to decimals that are one-dimensional. In our case, teachers urge the integration of theoretical insights in the intervention, whereas researchers often favor experimental designs driven by theoretical inquiries. The latter approach can be useful for isolating mechanisms, whereas the former has the potential to lead to greater learning gains overall.

Through codesign, teachers made critical contributions to a design which integrated these competing theories of rational number development, reinforcing *both* the bipartite relation (in line with Dewolf et al., 2015) and fraction magnitudes (in line with Siegler et al., 2015). In addition, teachers incorporated more fraction representations to promote knowledge transfer and stressed the importance of making clear connections between different formats

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(paper/pencil, board), which are in alignment with analogical reasoning principles (Richland, Zur, & Holyoak, 2007). Their intuitions, for example, that presenting students with different fraction representations side-by-side would be more appropriate for students "who need a challenge," is in line with our findings that students with higher prior knowledge benefited more from simultaneous presentations than students with lower prior knowledge. These alignments are further discussed in the context of research questions 1–3.

In addition, teachers' and students' ideas often addressed on-the-ground challenges common to school contexts, which, in turn, improved the overall usability of the intervention. For example, the FEMINEERS purposefully avoided using technologies like 3D printers with the goal of providing a design solution that could easily be implemented with minimal resources. Teachers confirmed that these design decisions would ease implementation and increase the game's flexibility to include more representations of fraction magnitudes.

Through a design narrative, we detailed our process of cocreating these activities with teachers and students. We hope this narrative can serve as a model for education and psychology researchers to partner with educators and students to create programs that are aligned with education, developmental, and cognitive science, and also takes advantage of educator and students' expertise in their own educational context for a product that is usable, effective, and fun.

4.2. RQ1: Does Fraction Ball/Bottle Caps Bonanza improve student outcomes compared to business as usual (including far transfer)?

4.2.1. Contributions of our findings

Similar to previous studies, Fraction Ball improved students' ability on the math they practiced during the games. In contrast to previous Fraction Ball studies, students were also able to transfer their rational number knowledge to fraction representations not presented during the intervention (e.g., fifths, tenths, or improper fractions).

We posit that this intervention was particularly potent because it integrates codesign and science of learning principles. More broadly, we theorize this approach is a promising avenue for designing usable and effective interventions (see Begolli et al., 2024). The next section summarizes how key principles from the science of learning (e.g., analogical reasoning and number lines) were integrated into our intervention, and we speculate on the influence this may have had on student learning.

4.2.2. Infusing science of learning principles into playful learning interventions

Fraction Ball and Bottle Caps Bonanza improved students' overall rational number (RQ1) and far transfer fraction skills by integrating science of learning principles with student and teacher expertise. It is likely that magnitude estimation skills were most impacted by modeling our designs on the number line as a key mental representation (Siegler et al., 2011). Further, we suspect that impacts on far transfer were a result of children comparing and contrasting two play contexts, multiple representation formats (court, board, and paper), and multiple fraction representations (fourths, eighths, thirds, and sixths) in three formats (court,

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bottle Cap boards, and tracking sheets) as predicted by analogical reasoning theory (Gick & Holyoak, 1983).

4.2.2.1. Conditions under which play improves far transfer: The board game afforded opportunities for students to make connections between mental representations and calculations on the court, life-size number line, and arithmetic sentences on paper. Theoretically, repeated practice across two contexts and multiple formats may lead reasoners to disregard surface commonalities and differences, instead attending to, aligning, and mapping between structural features that are common across these games and formats, thereby promoting schema building and far transfer (see Holyoak & Thagard, 1997; Gick & Holyoak, 1983). In alignment with analogical reasoning theory and teacher intuitions, we drew on the progressive alignment theory, purporting that reasoners should start by comparing representations that have the greatest commonalities (Kotovsky & Gentner, 1996). Thus, we chose proportionally equivalent, or "friendly," unlike denominators and presented fourths with eighths in one version of the game and thirds with sixths in another. Repeated practice with varied fraction magnitude representations arguably contributed to schema building and far transfer, as would be predicted by analogical reasoning principles.

We find that students who learned through traditional instruction (BAU) also experienced gains with more common fraction representations (e.g., fourths, thirds, sixths, and eighths), but did not transfer their knowledge to less common representations (e.g., ⅕ and 1/10).

4.2.2.2. Conditions under which play improves fraction knowledge: The impacts on overall fraction knowledge were smaller than in our previous Fraction Ball iterations (e.g., Begolli et al., 2024; Bustamante et al., 2022), which may be due to reduced dosage (4 vs. 12 sessions). The positive impacts seem to be driven by moderate effects on number line estimation, where we found the greatest learning gains, likely from repeated practice on the number line. Similar to the "Putting Fractions Together" intervention (Braithwaite & Siegler, 2021), the game design and rules build on children's previous knowledge with whole numbers. For example, a unit fraction is represented as the smallest point value in the game, with opportunities to generate the total sum of a score (quantity) by concatenating unit fractions only or by concatenating unit and nonunit fractions. For example, to add $\frac{1}{4}$ and $\frac{3}{4}$, students can start at 0 and add $\frac{1}{4}$, and then from $\frac{1}{4}$ count, three additional $\frac{1}{4}$ hatch marks along the number line to reach 1 whole. This is analogous to whole number counting and addition strategies.

The number lines, which served as scoreboards, had hatch marks for all fraction representations and whole numbers shown as integers, based on Siegler and Thompson's (2014) study. This design may have led students to treat fractions larger than 1 as mixed numbers. In turn, this may have influenced them to correctly add and compare fractions as well as integrate their understanding of whole number and fraction magnitudes (Siegler & Thompson, 2014).

Strategy choices and procedures seem to work in a cyclical fashion with conceptual understanding, such that conceptual knowledge serves to support procedural knowledge and vice versa (Rittle-Johnson, Siegler, & Alibali, 2001), which aligned with teachers' on-the-ground judgments. By design, in the Bottle Cap Bonanza, students wrote an arithmetic sentence each time they scored a point, and in Fraction Ball, they were expected to say the sentence out loud.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Drawing on the principle that physical objects are meant to embody core mathematical concepts, the board/court and scoreboard embody the number line as a physical representation of fraction magnitudes (see Link et al., 2013 for an example with whole numbers). Exploration prior to instruction (i.e., preparation for future learning; Schwartz et al., 2011) is embedded in the game mechanics, as well. Students are prompted to engage with fraction addition with like and unlike denominators, before a teacher steps in to support children's reasoning.

4.3. RQ2: Does the order of presenting representations within the combination of Fraction Ball and Bottle Caps Bonanza impact fraction learning? Specifically, does a simultaneous version (equivalent fraction representations simultaneously visible) improve fraction knowledge compared to a sequential version (equivalent fraction representations shown sequentially)?

We created two versions of the intervention, one that presented fractions sequentially and one simultaneously on the board and on the court. Within our team, as well as in our codesign sessions with teacher partners, we were conflicted as to which version of the intervention would lead to better outcomes. In both versions, fractions with "friendly" unlike denominators (e.g., $\frac{1}{4}$ and 2/8) are presented side-by-side on the number line to support the notion of equivalency between the representations. We did not find any evidence of differences between sequential and simultaneous presentations in overall fraction knowledge. Students in the simultaneous condition did, however, score higher than students in the sequential condition on addition problems with unlike denominators—one of the most challenging rational number concepts and a frequent source of errors (Braithwaite & Siegler, 2024). This difference was not statistically significant ($b = 0.21$, $p = .05$), but it is possible that the simultaneous side-by-side presentations provided students more practice adding "friendly" unlike denominators and more opportunities to draw relationships between these representations through processes of mapping and aligning, leading to deeper knowledge of addition with unlike denominators.

Our study presents another data point that playful learning activities lead to knowledge improvements compared to traditional instruction, with a caveat. We found suggestive evidence that student learning depends on an interaction between the kinds of representations embedded in the game environment and students' prior knowledge. These interactions are discussed next.

4.4. RQ3: Are the treatment impacts on students' fraction knowledge moderated by students' prior knowledge and grade level?

We found that students with higher prior knowledge benefited more from simultaneous presentations of fractions with "friendly" unlike denominators. In contrast, we found that students with lower prior knowledge may even be hindered in this condition; however, the difference was not statistically significant ($b = -0.12$, $p = .08$). These results are in line with prior studies and may reflect the challenges imposed on cognitive resources required to successfully learn by comparing and contrasting representations (Guo & Pang, 2011). Structural alignment and mapping places a significant burden on executive functions (Waltz et al., 2000),

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and higher prior knowledge may serve to offload executive function resources by making representations better understood and thus easier to hold in mind, enabling reasoners to focus their resources on similarities and differences that lead to successful mapping and, in turn, schema building. Conversely, students with low prior knowledge in this study may have experienced an overload of their executive functions, potentially explaining their lower impacts (see Begolli et al., 2018), though the difference was not statistically significant ($b = -0.12$, $p = .08$). Future studies should examine these mechanisms further. From a design perspective, it appears that we should perhaps begin with sequential presentations and then move on to simultaneous presentations in order to benefit all learners.

In contrast to prior knowledge, we did not find differences between grade levels, suggesting the Fraction Ball and Bottle Cap Bonanza games are appropriate across fourth and fifth grade, benefiting students of both levels. It may also reflect that grade level may not be the best predictor of students' prior knowledge and/or potential to learn.

4.5. Limitations

Our design, intervention, and evaluation processes provide a powerful method for developing educational programs that impact student learning and give insight into learning principles that lead to schema building and far transfer of fraction concepts. Mechanistically, we confirm prior evidence that simultaneous presentations benefit high prior knowledge students more than sequential presentations. In our follow-up work, we drew from these findings and prior work (Guo & Pang, 2011) to present comparisons first sequentially, then simultaneously. There are many other learning principles that are embedded within the intervention (e.g., number line training), arising from the science of learning and codesign. These combinations are perhaps the greatest strength of the program, contributing to potent student impacts, but also make it challenging to disentangle the mechanistic contributions of each principle.

In contrast to previous interventions (Begolli et al., 2024; Bustamante et al., 2020), the current study was fully experimenter-led, which puts into question its generalizability, namely, whether the same impacts and variations would be demonstrated, if teachers were to lead the intervention. Also, classroom management may pose another threat to generalizability because classes were split in half due to randomization, and smaller groups are typically easier to manage. Additionally, the fidelity data were not collected in this study; this represents a limitation, which restricts our ability to assess implementation consistency across classrooms. In our follow-up work, we have developed a more extensive version of the program that incorporates both Fraction Ball and Bottle Cap Bonanza as well as sequential and simultaneous presentations of fractions with unlike denominators, in a fully teacher-led study.

5. Conclusion

Our study presents a play-based intervention that demonstrates significant impacts on far transfer and fraction knowledge, specifically, compared to a business-as-usual math class. It also builds on our previous work, showing that play-based interventions can be a powerful tool

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for improving fraction knowledge. In line with previous literature in analogical reasoning, interacting with two play contexts as well as multiple formats and representations leads to schema building that supports knowledge transfer to novel fraction representations.

An important takeaway for designers is that the success or pitfall of play-based educational interventions depends on the strategic integration of science of learning principles and community expertise into the game design. Our data exemplify that the manner of presentation (sequential or simultaneous) can lead to differential effects depending on children's prior knowledge. Specifically, the simultaneous group showed higher posttest gains compared to the sequential group, though this difference did not reach the conventional threshold for statistical significance and may not generalize beyond this sample ($b = 0.21$, $p = .05$). Conversely, students with higher prior knowledge benefited more from simultaneous presentations.

In addition, teachers provided important insights into the challenges their students face with fraction concepts and provided potential solutions that aligned with science of learning principles from analogical reasoning and number line literature. Students provided solutions that aligned with playful learning principles (e.g., engaging, socially interactive, etc.) that afforded a low-cost, flexible, and highly scalable play environment, which embedded learning principles found in previous iterations of Fraction Ball by developing the Bottle Cap Bonanza game. Thus, through codesign, we found that researchers, teachers, and students' expertise aligned within a playful learning context that resulted in a potent intervention that improved what has traditionally been one of the most challenging skills to tackle—knowledge transfer.

References

Ahn, J., Campos, F., Hays, M., & DiGiacomo, D. (2019). Designing in context: Reaching beyond usability in learning analytics dashboard design. *Journal of Learning Analytics*, *6*(2), 70–85. https://doi.org/10.18608/jla.2019.62.5

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Anderson, R. D., & Helms, J. V. (2001). The ideal of standards and the reality of schools: Needed research. *Journal of Research in Science Teaching*, *38*(1), 3–16. https://doi.org/10.1002/1098-2736(200101)38:1⟨3::AID-TEA2⟩3.0.CO;2-V

Anderson-Coto, M. J., Salazar, J., Lopez, J. L. S., Sedas, R. M., Campos, F., Bustamante, A. S., & Ahn, J. (2024). Towards culturally sustaining design: Centering community's voices for learning through participatory design. *International Journal of Child-Computer Interaction*, *39*, 100621. https://doi.org/10.1016/j.ijcci.2023.100621

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

Barbieri, C. A., Rodrigues, J., Dyson, N., & Jordan, N. C. (2020). Improving fraction understanding in sixth graders with mathematics difficulties: Effects of a number line approach combined with cognitive learning strategies. *Journal of Educational Psychology*, *112*(3), 628.

Begolli, K. N., Bermudez, V. N., Lawrence, L., Acevedo-Farag, L. M., Valdez, S. V., Santana, E., Acevedo-Farag, L., Bailey, D., Rhodes, K., Richland, L. E., Ahn, J., & Bustamante, A. S. (2024). Incorporating design based implementation research with a randomized controlled trial to develop and evaluate the efficacy of playful rational number learning. *Contemporary Educational Psychology*, *78*, 102296.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Begolli, K. N., Richland, L. E., Jaeggi, S. M., Lyons, E. M., Klostermann, E. C., & Matlen, B. J. (2018). Executive function in learning mathematics by comparison: Incorporating everyday classrooms into the science of learning. *Thinking & Reasoning*, *24*(2), 280–313.

Begolli, K. N., & Richland, L. E. (2016). Teaching mathematics by comparison: Analog visibility as a double-edged sword. *Journal of Educational Psychology*, *108*(2), 194–213. https://doi.org/10.1037/edu0000056

Belgrave, A. B., Bermudez, V. N., Chew, P., Lin, Y., Ahn, J., & Bustamante, A. S. (2022). Using a participatory design approach for co-creating culturally situated STEM enrichment activities. *Journal of Applied Developmental Psychology*, *82*, 101451.

Bell, P., Hoadley, C. M., & Linn, M. C. (2013). Design-based research in education. In *Internet environments for science education* (pp. 101–114). Routledge.

Bermudez, V. N., Salazar, J., Garcia, L., Ochoa, K. D., Pesch, A., Roldan, W., … & Bustamante, A. S. (2023). Designing culturally situated playful environments for early STEM learning with a Latine community. *Early Childhood Research Quarterly*, *65*, 205–216.

Bødker, S., Dindler, C., Iversen, O. S., & Smith, R. C. (2022). What is participatory design?. In *Participatory design. Synthesis Lectures on Human-Centered Informatics*. Cham. https://doi.org/10.1007/978-3-031-02235-7_2

Booth, J. L., McGinn, K. M., Barbieri, C., Begolli, K. N., Chang, B., Miller-Cotto, D., Young, L. K., & Davenport, J. L. (2017). Evidence for cognitive science principles that impact learning in mathematics. In D. C. Geary, D. B. Berch, R. Ochsendorf, & K. M. Koepke (Eds), *Acquisition of complex arithmetic skills and higher-order mathematics concepts* (pp. 297–325). Elsevier.

Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology*, *37*(4), 247–253.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*(1), 189–201. https://doi.org/10.1037/0012-1649.41.6.189

Bottge, B. A., Ma, X., Gassaway, L., Toland, M. D., Butler, M., & Cho, S.-J. (2014). Effects of blended instructional models on math performance. *Exceptional Children*, *80*(4), 423–437.

Braithwaite, D. W., & Siegler, R. S. (2021). Putting fractions together. *Journal of Educational Psychology*, *113*(3), 556.

Braithwaite, D. W., & Siegler, R. S. (2024). A unified model of arithmetic with whole numbers, fractions, and decimals. *Psychological Review*, *131*(2), 431

Braithwaite, D. W., Tian, J., & Siegler, R. S. (2018). Do children understand fraction addition? *Developmental Science*, *21*(4), e12601.

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*(1), 61–100. https://doi.org/10.3102/0091732X024001061

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, *2*(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Bustamante, A. S., Begolli, K. N., Alvarez-Vargas, D., Bailey, D. H., & Richland, L. E. (2022). Fraction ball: Playful and physically active fraction and decimal learning. *Journal of Educational Psychology*, *114*(6), 1307.

Bustamante, A. S., Schlesinger, M., Begolli, K. N., Golinkoff, R. M., Shahidi, N., Zonji, S., Riesen, C., Evans, N., & Hirsh-Pasek, K. (2020). More than just a game: Transforming social interaction and STEM play with Parkopolis. *Developmental Psychology*, *56*(6), 1041.

Byrnes, J. P., & Wasik, B. A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology*, *27*(5), 777.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13. https://doi.org/10.3102/0013189X032001009

Common Core State Standards Initiative. (2012). *Common core state standards for mathematics*. Common Core State Standards Initiative.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Dewolf, M., Bassok, M., & Holyoak, K. J. (2015). Conceptual structure and the procedural affordances of rational numbers: Relational reasoning with fractions and decimals. *Journal of Experimental Psychology: General*, *144*(1), 127–150.

DiSalvo, B., Yip, J., Bonsignore, E., & Carl, D. (2017). *Participatory design for learning*. Routledge.

Dyson, N. I., Jordan, N. C., Rodrigues, J., Barbieri, C., & Rinne, L. (2020). A fraction sense intervention for sixth graders with or at risk for mathematics difficulties. *Remedial and Special Education*, *41*(4), 244–254.

Else, M. J., & Clement, J. (2003). *Should different types of analogies be treated differently in instruction?*

Fazio, L. K., Kennedy, C. A., & Siegler, R. S. (2016). Improving children's knowledge of fraction magnitudes. *PLoS One*. *11*(10), e0165243.

Fazio, L. K., & Siegler, R. S. (2010). Teaching fractions. *International Academy of Education*.

Fischer, U., Moeller, K., Bientzle, M., Cress, U., & Nuerk, H.-C. (2011). Sensori-motor spatial training of number magnitude representation. *Psychonomic Bulletin & Review*, *18*(1), 177–183.

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., Jordan, N. C., Siegler, R., Gersten, R., & Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology*, *105*(3), 683–700. https://doi.org/10.1037/a0032446

Galleguillos, M. L. R., & Coskun, A. (2020). How do I matter? A review of the participatory design practice with less privileged participants. In *Proceedings of the 16th Participatory Design Conference*, Vol. 1, 137–147. https://doi.org/10.1145/3385010.3385018

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. https://doi.org/10.1016/S0364-0213(83)80009-3

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences*, *14*(1), 69–110.

Gooch, D., Barker, M., Hudson, L., Kelly, R., Kortuem, G., Linden, J. V. D., … & Walton, C. (2018). Amplifying quiet voices: Challenges and opportunities for participatory design at an urban scale. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *25*(1), 1–34. https://doi.org/10.1145/3139398

Gray, M. E., & Holyoak, K. J. (2021). Teaching by analogy: From theory to practice. *Mind, Brain, and Education*, *15*(3), 250–263.

Guo, S., Bailey, D. H., Rhodes, K., Begolli, K. N., Bermudez, V. N., Lawrence, L., Alvarez-Vargas, D., Acevedo-Farag, L. M., Ahn, J., Bustamante, A. S., & Richland, L. E. (2024). Buckets of fun: Impacts of fraction ball activities on students' math-related emotions. *Journal of Applied Developmental Psychology*, *92*, 101645.

Guo, J., & Pang, M. (2011). Learning a mathematical concept from comparing examples: The importance of variation and prior knowledge. *European Journal of Psychology of Education*, *26*(4), 495–525. https://doi.org/10.1007/s10212-011-0060-y

Hamdan, N., & Gunderson, E. A. (2017). The number line is a critical spatial-numerical representation: Evidence from a fraction intervention. *Developmental Psychology*, *53*(3), 587–587.

Handel, M. J. (2016). What do people do at work? *Journal for Labour Market Research*, *49*(2), 177–197.

Hassinger-Das, B., Bustamante, A., Hirsh-Pasek, K., & Golinkoff, R. M. (2018). Learning landscapes: Playing the way to learning and engagement in public spaces. *Education Sciences*, *8*(2), 74. https://doi.org/10.3390/educsci8020074

Hattikudur, S., & Alibali, M. W. (2010). Learning about the equal sign: Does comparing with inequality symbols help? *Journal of Experimental Child Psychology*, *107*(1), 15–30. https://doi.org/10.1016/j.jecp.2010.03.004

Hecht, S. A. (1998). Toward an information-processing account of individual differences in fraction skills. *Journal of Educational Psychology*, *90*(3), 545–559. https://doi.org/10.1037/0022-0663.90.3.545

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological science in the public interest*, *16*(1), 3–34.

Hoadley, C. P. (2023, January). Creating context: Design-based research in creating and understanding CSCL. In *Computer support for collaborative learning* (pp. 453–462). Routledge.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Holyoak, K. J., & Thagard, P. (1997). The analogical mind. *The American Psychologist*, *52*(1), 35–44. https://doi.org/10.1037/0003-066X.52.1.35

IJzerman, H., Lewis Jr, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, *4*(11), 1092–1094. https://doi.org/10.1038/s41562-020-00990-w

Koedinger, K. R., Corbett, A., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, *36*, 757–798. https://doi.org/10.1111/j.1551-6709.2012.01245.x

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*, 2797–2822. https://doi.org/10.2307/1131753

Lawrence, L., Szura, J. W., Begolli, K. N., Minko, J., Bustamante, A., & Ahn, J. (2022). Tiered levels of engagement for teacher partnerships. In L. Wentworth, C. Conaway, S. Shewchuk, & P. Arce-Trigatti (Eds.), *RPP Brokers Handbook, V. 2: A Guide to Brokering in Education Research-Practice Partnerships*. Houstin, TX: National Network of Education Research-Practice Partnerships (NNERPP).

Link, T., Moeller, K., Huber, S., Fischer, U., & Nuerk, H. (2013). Trends in neuroscience and education walk the number line – An embodied training of numerical concepts. *Trends in Neuroscience and Education*, *2*(2), 74–84. https://doi.org/10.1016/j.tine.2013.06.005

Mazzocco, M. M. M., & Devlin, K. T. (2008). Parts and 'holes': Gaps in rational number sense among children with vs. without mathematical learning disabilities. *Developmental Science*, *11*(5), 681–691.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, *16*(5), 519–533.

Moser, A., Zimmermann, L., Dickerson, K., Grenell, A., Barr, R., & Gerhardstein, P. (2015). They can interact, but can they learn? Toddlers' transfer learning from touchscreens and television. *Journal of experimental child psychology*, *137*, 137–155.

National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. In *Foundations* (Vol. 37). https://doi.org/10.3102/0013189X08329195

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. National Academies Press.

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.

Penuel, W. R., & Fishman, B. J. (2012). Large-scale science education intervention research we can use. *Journal of Research in Science Teaching*, *49*(3), 281–304. https://doi.org/10.1002/tea.21001

Penuel, W. R., & Gallagher, D. J. (2017). *Creating research practice partnerships in education*.

Pomerance, L., Greenberg, J., & Walsh, K. (2016). *Learning about Learning: What Every New Teacher Needs to Know*. National Council on Teacher Quality.

Richland, L. E., & Begolli, K. N. (2016). Analogy and higher-order thinking: Learning mathematics as an example. *Policy Insights from the Behavioral and Brain Sciences*, *4*, 1–18.

Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the conceptual structure of mathematics. *Educational Psychologist*, *47*(3), 189–203. https://doi.org/10.1080/00461520.2012.667065

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science*, *316*(5828), 1128–1128.

Richland, L. E., Begolli, K. N., Simms, N., Frausel, R. R., & Lyons, E. A. (2017). Supporting mathematical discussions: The roles of comparison and cognitive load. *Educational Psychology Review*, *29*(1), 41–53.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*(2), 346–362. https://doi.org/10.1037/0022-0663.93.2.346

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, *101*(4), 836–852. https://doi.org/10.1037/a0016026

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, *27*(4), 587–597.

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 456–468. https://doi.org/10.1037/0278-7393.15.3.456

Sanders, E. B. N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *Co-design*, *4*(i), 5–18. https://doi.org/10.1080/15710880701875068

Schlesinger, M., Hassinger-Das, B., Zosh, J. M., Sawyer, J., Evans, N., & Hirsh-Pasek, K. (2020). Cognitive behavioral science behind the value of play: Leveraging everyday experiences to promote play, learning, and positive interactions. *Journal of Infant, Child, and Adolescent Psychotherapy*, *19*(2), 202–216.

Schumacher, R. F., Jayanthi, M., Gersten, R., Dimino, J., Spallone, S., & Haymond, K. S. (2018). Using the number line to promote understanding of fractions for struggling fifth graders: A formative pilot study. *Learning Disabilities Research & Practice*, *33*(4), 192–206.

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. *Transfer of learning from a modern multidisciplinary perspective*, *3*(1), 1–51.

Schwartz, D. L., Chase, C. C., Oppezzo, M. a., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759–775. https://doi.org/10.1037/a0025140

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, *11*(5), 655–661.

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

Siegler, R. S., & Thompson, C. A. (2014). Numerical landmarks are useful—except when they're not. *Journal of experimental child psychology*, *120*, 39–58.

Simonsen, J., & Robertson, T. (2013). *Routledge international handbook of participatory design* (Vol. 711). New York: Routledge.

Star, J. R., Rittle-Johnson, B., & Durkin, K. (2016). Comparison and explanation of multiple strategies: One example of a small step forward for improving mathematics education. *Policy Insights from the Behavioral and Brain Sciences*, *3*(2), 151–159. https://doi.org/10.1177/2372732216655543

Thompson, C. A., & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development*, *81*(6), 1768–1786. https://doi.org/10.1111/j.1467-8624.2010.01509.x

Waltz, J. A., Lau, A., Grewal, S. K., Holyoak, K. J., Lad, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition*, *28*(7), 1205–1212.

Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A. S., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *38*(5), 529–552.

Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., Solis, S. L., & Whitebread, D. (2018). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in Psychology*, *9*(AUG), 1–12. https://doi.org/10.3389/fpsyg.2018.01124

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary information

15516709, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cogs.70168 by University Of California Irvine Science Library Serials, Wiley Online Library on [14/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License