







# Perceiving Precedence: Order of Operations Errors Are Predicted by Perception of Equivalent Expressions

Jeffrey Kramer Bye<sup>1</sup> , Jenny Yun-Chen Chan<sup>2</sup> , Avery H. Closser<sup>3</sup> , Ji-Eun Lee<sup>4</sup> , Stacy T. Shaw<sup>4</sup> ,  
Erin R. Ottmar<sup>4</sup> 

[1] Department of Psychology, California State University, Dominguez Hills, Carson, CA, USA. [2] Department of Early Childhood Education, The Education University of Hong Kong, Tai Po, Hong Kong, SAR, China. [3] School of Teaching and Learning, University of Florida, Gainesville, FL, USA. [4] Department of Social Science & Policy Studies, Worcester Polytechnic Institute, Worcester, MA, USA.

Journal of Numerical Cognition, 2024, Vol. 10, Article e14103, <https://doi.org/10.5964/jnc.14103>

Received: 2024-03-03 • Accepted: 2024-09-20 • Published (VoR): 2024-12-04

Handling Editor: Joonkoo Park, University of Massachusetts Amherst, Amherst, MA, USA

Corresponding Author: Jeffrey Kramer Bye, Department of Psychology, 1000 E. Victoria Street, Carson, CA, USA, 90747-0001. E-mail: [jkbye@csudh.edu](mailto:jkbye@csudh.edu)

Supplementary Materials: Code, Data, Preregistration [see Index of Supplementary Materials]



## Abstract

Students often perform arithmetic using rigid problem-solving strategies that involve left-to-right-calculations. However, as students progress from arithmetic to algebra, entrenchment in rigid problem-solving strategies can negatively impact performance as students experience varied problem representations that sometimes conflict with the order of precedence (the order of operations). Research has shown that the syntactic structure of problems, and students' perceptual processes, are involved in mathematics performance and developing fluency with precedence. We examined 837 U.S. middle schoolers' propensity for precedence errors on six problems in an online mathematics game. We included an algebra knowledge assessment, math anxiety measure, and a perceptual math equivalence task measuring quick detection of equivalent expressions as predictors of students' precedence errors. We found that students made more precedence errors when the leftmost operation was invalid (addition followed by multiplication). Individual difference analyses revealed that students varied in propensity for precedence errors, which was better predicted by students' performance on the perceptual math equivalence task than by their algebra knowledge or math anxiety. Students' performance on the perceptual task and interactive game provide rich insights into their real-time understanding of precedence and the role of perceptual processes in equation solving.

## Keywords

arithmetic, precedence, order of operations, perceptual learning, educational technology, math equivalence, individual differences



### Highlights

- We examined 837 U.S. middle school students' performance on pretest measures of algebra knowledge, math anxiety, and perception of equivalence, as well as how they solved problems in an online mathematics game.
- Students were over three times as likely to make errors in precedence (order of operations) when the leftmost operation could not be computed first (e.g.,  $8 + 4 \times 4 \times 8$ ) than when it could (e.g.,  $3 \times 7 + 7 + 3$ ).
- These precedence errors varied considerably across students and were best explained by students' ability to perceive equivalent mathematical expressions, e.g., perceiving  $2(3y - 6)$  as equivalent to  $-2 \times 6 + 2 \times 3y$ .
- These results extend prior research on perceptual learning in mathematics to suggest that students' ability to perceive expressions as equivalent is an important aspect of perceiving precedence, over and above other measures of algebra knowledge.

The order of operations (rules of precedence; hereafter precedence) involves performing higher-order arithmetic operations (e.g., multiplication) before lower-order operations (e.g., addition). For example, in  $5 + 2 \times 3$ , the 2 and 3 must first be multiplied before their product (i.e., 6) can be added to 5. However, students in many cultures learn precedence after years of experience solving problems left-to-right, so even middle schoolers may attempt to add 5 and 2 first, yielding an incorrect result (Banerjee & Subramaniam, 2005; Blando et al., 1989; Gunnarsson et al., 2016; Kieran, 1979). These precedence errors are common mistakes in arithmetic and algebra, and may reflect students' struggles in breaking away from left-to-right solving, which is essential for transitioning from arithmetic to algebra (Norton & Cooper, 2001).

Recent work has highlighted the importance of perceptual learning to mathematics, i.e., the ability to fluently extract information about structure, such as precedence, from mathematical expressions (Kellman et al., 2010; Kellman & Massey, 2013). Specifically, perception of notation structure is associated with algebraic reasoning skills among adults (Marghetis et al., 2016); additionally, such perception can be trained in middle and high school students (via perceptual learning modules), which leads to improved learning outcomes even at a delay (Kellman et al., 2010). Collectively, this work suggests that through practice, students learn to extract mathematical information (e.g., that  $2 \times 3$  takes precedence in  $5 + 2 \times 3$ ) by fine-tuning their visual perceptual processes; this extracted structure may then serve procedural and conceptual aspects of mathematical reasoning (e.g., inhibiting a left-to-right habit to compute  $2 \times 3$  first; Kellman et al., 2010). However, it is less clear how perceptual fluency is associated with problem-solving skills (and errors) in younger students.

Here, we aimed to better understand factors contributing to precedence errors and potential ways to help students overcome this challenge. We examined middle schoolers' problem-solving performance as they simplified six arithmetic problems within an online game and analyzed how the composition of problems (i.e., a syntactic structure of equations) related to students' propensity to make errors. We also examined potential student characteristics that may explain who is most likely to make precedence errors. We included a measure of students' perception of algebraic equivalence (inspired by work in perceptual learning; Kellman et al., 2010; Marghetis et al., 2016) to test its relation with students' precedence error propensity above and beyond known predictors of students' mathematics performance—algebraic knowledge (Spielhagen, 2006) and math anxiety (Foley et al., 2017). Together, these analyses reveal how the presentation of problems in online settings impact students' problem-solving behavior and performance and how the composition of problems interacts with other student characteristics.

### Entrenchment in Left-to-Right Solving Strategies

In many cultures, students commonly solve problems from left to right even when this violates precedence (Banerjee & Subramaniam, 2005; Blando et al., 1989; Gunnarsson et al., 2016; Norton & Cooper, 2001). This left-to-right approach is likely a procedural routine that students learned through narrow experience with arithmetic and became entrenched in their (mis)understanding of mathematics (McNeil, 2014). Although performing left-to-right calculations is a common strategy that generalizes across many arithmetic problems (e.g.,  $4 + 3 - 2 = \underline{\quad}$ ), it fails to generalize to multiple problem representations that occur in pre-algebra tasks (e.g.,  $4 + 3 = \underline{\quad} + 2$ ). Accordingly, students struggle to solve pre-algebra equations when the syntactic structures of equations do not follow the well-practiced “operations on the left and the

answer on the right” format. In this example, many elementary school students write 7 (i.e.,  $4 + 3$ ) as the answer, ignoring the  $+2$  on the right of the empty space (McNeil & Alibali, 2005). Similarly, after solving arithmetic problems (e.g.,  $6 + 8 = \underline{\quad}$ ), adults were *less* likely to solve pre-algebra equations correctly, likely because their early procedural habits of solving left-to-right were activated (McNeil et al., 2010). This strict adherence to left-to-right solving reflects students’ misconceptions about equivalence and negatively predicts their later algebra readiness (Hornburg et al., 2022).

Entrenchment in left-to-right problem solving causes immediate and long-term challenges for students. First, it leads to incorrect answers for expressions with mixed-precedence structures (e.g.,  $2 + 3 \times 4$ ) unless the higher-precedence operations happen to be on the left (e.g.,  $3 \times 4 + 2$ ). Second, even when left-to-right solving still yields the correct answer (e.g.,  $3 \times 4 + 5 - 5$ ), it may not be the most efficient strategy. Flexible problem solving—the ability to generate multiple strategies and to identify efficient strategies (Hong et al., 2023)—involves noticing algebraic structures and leveraging these structures to simplify problems (e.g., by performing subtraction first, addition is negated, removing a step). This flexibility is essential for both arithmetic and algebra (Star & Rittle-Johnson, 2008; Yu et al., 2018), but it requires a strong structural understanding of precedence (e.g., both multiplication and subtraction are valid first actions in the previous example).

Although precedence emerges as an issue in arithmetic, it is especially crucial to understand to succeed in algebra and advanced mathematics subjects. Students must understand that adhering to precedence does not always mean performing calculations sequentially based on arbitrary conventions such as PEMDAS (Parentheses, Exponents, Multiplication, Division, Addition, Subtraction) or BODMAS (Brackets, Orders, Division, Multiplication, Addition, Subtraction). They are often taught to memorize acronyms like PEMDAS or BODMAS which may exacerbate their difficulties with flexible problem solving (Linchevski & Livneh, 1999). For example, overreliance on the order of terms in PEMDAS can lead to multiplication being performed before division, such that  $10 \div 5 \times 2$  is calculated as  $10 \div 10$  (i.e., 1) instead of  $2 \times 2$  (i.e., 4). Such misconceptions have even been commonly found among prospective teachers (Glidden, 2008; Zazkis & Rouleau, 2018). How do students move beyond a mnemonic approach like PEMDAS/BODMAS for more flexible problem solving in arithmetic, setting up a transition to algebra? Based on the foundation of perceptual learning (Kellman et al., 2010; Marghetis et al., 2016), we propose that *perceiving* the hierarchical structure of symbols within problems may be one mechanism behind students’ fluency and flexibility in problem solving that emerges from understanding precedence and practicing mixed-precedence problems solving.

## Perceptual Effects in Mathematical Structure

Students’ understanding of precedence is reflected in their perception of aesthetic features (e.g., color, spacing) and syntactic structures of notation (e.g., symbol arrangement). Perceptual learning theory states that individuals learn to leverage this sensory information from the environment to make accurate, efficient decisions (Gibson, 1969; see Szokolszky et al., 2019 for a review). Over time, individuals hone the ability to do so more effectively. Leveraging perceptual processes in this way is likely an evolutionary development that arose for general decision making and has been *exapted* (i.e., coopted from its original adaptive purpose; Gould, 1991) to advanced problem solving, even in the abstract notation and syntax of mathematics (Goldstone et al., 2017; Kellman & Massey, 2013).

Visual perceptual cues that are explicitly manipulated in math notation are often modeled after the Gestalt principles of grouping (Palmer, 1992; see Wagemans et al., 2012 for a review), which have been shown to influence students’ mathematical reasoning. For example, extensive research has replicated the effect of physical spacing between math symbols on secondary and college students’ immediate performance (e.g., Harrison et al., 2020; Jiang et al., 2014; Landy & Goldstone, 2007, 2010; Rivera & Garrigan, 2016). Specifically, students perform better when the spacing between symbols creates perceptual groupings aligned with precedence (e.g.,  $4 \times 5 + 6$ ) and worse when the spacing is not aligned with precedence (e.g.,  $4 \times 5 + 6$ ). Furthermore, twelve-year-olds relied *more* on the spacing between symbols and made more errors when the spacing was incongruent to precedence (e.g.,  $3 \times 4 + 5 - 5$ ) compared to eight-year-olds (Braithwaite et al., 2016), suggesting the potential increase in leveraging perceptual cues as students gain fluency in precedence (albeit inappropriately when viewing misleading, incongruently-spaced problems). Together, these findings suggest that students’ problem-solving performance may partly depend on these explicitly manipulated problem features as well as students’ perceptual fluency in grouping expressions around the higher-order operations.

## Contextual Effects of Syntactic Structure on Problem Solving Performance

Beyond these explicit manipulations to problems, even the syntactic structure of math notation can implicitly create visual perceptual cues, mimicking some Gestalt principles of grouping. In particular, the arrangement of symbols (Lee et al., 2022), symbol choice (Chan et al., 2022), and use of superfluous brackets in math problems (Egorova et al., 2024; Ngo et al., 2023) all impact students' problem-solving performance. As an example, adults search for, and attend to, higher-order operators (e.g., multiplication) more quickly than lower-order operators (e.g., addition; Landy et al., 2008), demonstrating their fluency in precedence and tendency to prioritize higher-order operators. However, even with this developed attention to higher-order operators, the position of higher-order operators in a mathematical expression still impacts students' performance. Undergraduate students are faster at simplifying expressions with the higher-order operator on the left ( $4 \times 5 + 6$ ) vs. right ( $6 + 4 \times 5$ ; Closser et al., 2023, 2024). Similarly, fifth- to seventh-grade students are more accurate and efficient at solving math problems when the higher-order operator is on the left as opposed to in the center (e.g.,  $6 + 4 \times 5 - 2$ ) or on the right side (Ngo et al., 2023), suggesting the influence of the syntactic structure on students' problem-solving performance. In summary, this research provides examples of how individuals find and leverage perceptual cues that are inherent to the composition of mathematical notation.

## Individuals' Perception of Algebraic Equivalence

Given that adults and experts rely on perceptual cues for accurate and efficient problem solving across contexts (e.g., chess: Chase & Simon, 1973; math: Kellman & Massey, 2013), perceptual fluency may be critical for algebraic problem solving. In algebra, concatenation represents multiplication (e.g.,  $3y = 3 \times y$ ) and variables cannot be computed directly (i.e., left-to-right computation) but must instead be 'isolated' via inverse operations. For example, Marghetis and colleagues (2016) designed a laboratory task to examine adults' perception of algebraic equivalence and how this perception is associated with attention to hierarchical structure of precedence. Specifically, they presented two algebraic expressions and asked participants whether the two expressions were equivalent or not equivalent. They found that adults who were more accurate at identifying equivalent expressions were also more likely to perceive subgroups of expressions following the order of precedence (i.e., better at detecting color changes within  $[a*b+c*d]$  vs. between  $[a*b+c*d]$  subgroups of expressions).

Similarly, Kellman and colleagues (2010) have designed tasks to train algebra students to recognize algebraic transformations through matching equivalent expressions that can be obtained by algebraic transformation (e.g.,  $2x + 3 = 5$  and  $2x = 5 - 3$ ). They argue that perceptual learning—"improvement in information extraction as a result of practice" (p. 286)—is relevant to selecting information and extracting structures, crucial to learning and developing expertise, and is theoretically as well as empirically distinct from conceptual or procedural learning. In their training of perceptual learning modules, students were presented a target equation with several options of alternative equations, and were asked to select the equation that could be obtained by a legal algebraic transformation of the target equation. The training led to improved fluency in equation solving among eighth and ninth grade students. The team also extended this work to the mapping across multiple representations (e.g., graphs and equations) and linear measurements, providing evidence for the role of perception in mathematical learning.

As noted by Star et al. (2015), the 2011 National Assessment of Educational Progress (NAEP) data demonstrated that less than 60% of US eighth-graders could identify which of 5 equations is not equivalent to  $n + 18 = 23$ . In a more recent and more difficult item, 2022 NAEP data revealed that *only* 6% of US eighth-graders could identify which 3 of 5 equations are equivalent to  $(2x - 4)/2 - (2 - x)$  (NAEP, 2022). Thus, it is important to develop measures of identifying equivalent expressions and equations, not only because this is an important (and difficult) task for students, but also because of the growing evidence that perceptual learning can help students to perceive equivalence. Grounded in the research on perceptual learning (Goldstone et al., 2017; Kellman & Massey, 2013; Kellman et al., 2010; Marghetis et al., 2016), we developed the Perceptual Math Equivalence Task (PMET) to measure students' perception of algebraic structure, specifically equivalence, and to understand how this perception impacts students' equation solving performance in authentic K-12 learning environments.

## Algebra Knowledge and Math Anxiety

The transition from arithmetic to algebra requires a fluent understanding of precedence, or what [Linchevski and Livneh \(1999\)](#) call ‘structure sense’. Precedence becomes increasingly important during this transition, and the differences between algebraic and arithmetic structures sometimes leads students to struggle with the ‘inverted’ order (e.g., in  $2x + 4 = 20$ , you must address the lower-precedence addition first, by subtracting 4). Prior studies have indicated that students’ understanding of precedence is a precondition for moving from arithmetic to pre-algebraic to algebraic problem solving ([Linchevski & Livneh, 1999](#); [Pillay et al., 1998](#); [Spitzer & Moeller, 2022](#)). Given the close association between precedence understanding and algebraic knowledge, we account for students’ performance on an algebra knowledge assessment ([Star et al., 2015](#)) when testing the effect of perception of algebraic equivalence on students’ propensity for precedence errors.

Another important construct that may affect precedence errors in students’ mathematical experiences is math anxiety ([Ramirez et al., 2018](#)). Specifically, there are two predominant accounts of math anxiety which both predict that students with higher math anxiety may be more likely to make mistakes such as precedence errors. The reduced competency account of math anxiety ([Maloney, 2016](#)) argues that poorer math skills contribute to math anxiety, and thus, higher math anxious students are more likely to make errors. The disruption account ([Ashcraft & Kirk, 2001](#)) instead argues that math anxiety hinders working memory resources, possibly making it more challenging to inhibit prepotent actions. Recent evidence suggests that both accounts may be partially correct in explaining people’s math errors ([Daker et al., 2023](#)). However, the extent to which math anxiety may contribute to this specific error is not clear, but worth exploration.

## The Present Study

In the present study, we systematically varied the syntactic structure of arithmetic expressions to examine middle school students’ propensity to make precedence errors. In particular, we examined whether students tend to implement left-to-right problem-solving strategies even when doing so violates the order of precedence. Different from traditional laboratory or pen-and-paper tasks that are often used in prior experimental research, we examined students’ precedence errors in an online mathematics game environment that encourages exploration and discovery of math relations. Furthermore, prior to the game, we provided students with pretest measures, including the Perceptual Math Equivalence Task (PMET) to measure students’ perception of algebraic equivalence (i.e., abilities to quickly perceive and identify equivalent expressions); we then examined the influences of this measure and others on students’ propensity for making errors in the order of operations (“precedence errors”). Through this approach, we explored how both problem-level characteristics and student-level differences influence problem-solving strategies.

Specifically, we examined students’ problem-solving processes on a select group of six target problems that varied in the position of the higher-order operator ([Table 1](#)); we analyzed under what conditions students were more likely to attempt to perform an operation for the lower-order operator (addition) prior to a higher-order operator (multiplication) when doing so violates the rules of precedence. In a smaller-scale pilot study with identical target problems but a separate sample, we found that students varied in propensity for precedence errors at the beginning of the problem, which was better predicted by their perception of equivalence than algebraic knowledge ([Bye et al., 2022](#)). Extending our pilot study, we preregistered three research questions, hypotheses, and their analytic approach on the Open Science Framework (OSF) prior to processing the log data for precedence errors (see [Bye et al., 2023S](#)).

### **RQ1: Does the frequency of students’ initial precedence errors vary depending on the perceptual features (e.g., position of the operator) of the problems?**

In RQ1, we asked whether students made more precedence errors at the beginning of the problem as a function of the problem structure. Based on results from our pilot study ([Bye et al., 2022](#)) and prior research on order of operations, we hypothesized that students in aggregate would make more precedence errors when performing the leftmost operation initially would be invalid (e.g.,  $8 + 4 \times 4 \times 8$ ) than when it would be valid (e.g.,  $3 \times 7 + 7 + 3$ ).

**RQ2: Do students' algebra knowledge, math anxiety, or perception of equivalence predict their precedence error frequency on their first step of problem solving?**

In RQ2, we assessed to what extent students' propensity in making these precedence errors was explained by their algebra knowledge, math anxiety, or perception of equivalence (i.e., identifying when expressions/equations are equivalent). Again, based on our pilot study as well as prior research in perceptual learning, we hypothesized that students' perception of equivalent expressions would be a better predictor of precedence errors than their algebra knowledge or math anxiety.

**RQ3: Do students' precedence errors persist in their subsequent second step of the problem-solving process?**

In RQ3, we examined whether students persisted in making precedence errors even after their first valid operation. Here, we applied the same analytic framework as RQ1 and RQ2 to students' precedence errors in their *second* step of the problem-solving process. That is, after a valid move, if it is now invalid to perform the leftmost action in the transformed expression (e.g.,  $6 + 3 \times 2$ ), will students be more likely to produce a precedence error than if it is valid (e.g.,  $6 \times 6 + 1$ )? As far as we know, no study has looked at precedence errors at this level of detail or in this specific aspect of multistep problem solving. We hypothesized for RQ3 that the problem-level features (as in RQ1 for the first step) and individual differences (as in RQ2 for the first step) would both remain predictive of second-step precedence errors.

RQs 1 and 2 are adapted from our pilot study (Bye et al., 2022) with a smaller sample size ( $N = 217$ ), thus these RQs in the present study also assess whether our original results are replicated in this new, larger sample ( $N = 837$ ).

## Method

### Participants

Participants were drawn from a larger randomized control trial (RCT) conducted in 11 middle schools (10 in-person schools and one virtual academy) in the Southeastern U.S. testing the efficacy of four math technology interventions during the 2020-2021 academic year (see Decker-Woodrow et al., 2023, for more detailed information about the larger study). Among the 4,092 seventh grade students (i.e., ages 12-13) who participated in the larger study, 1,649 were assigned to a math game condition (From Here to There! [FH2T]), including students in resource settings. An additional 37 students moved into the participating district mid-year and were assigned to the FH2T condition. Because we had access to these students' pretests and logs for the focal study problems, we also included them in the current study pool. One additional participant was removed from analyses due to an implausible time recorded on the PMET. A total of 837 of the FH2T students completed the pretest and solved all six focal problems so only these students were included in the analyses for the current study. This final analytic sample ( $N = 837$ ) was much smaller than the original FH2T sample ( $N = 1,649$ ) in the larger RCT because (a) two schools opted out of the RCT due to instructional concerns related to COVID-19 ( $n = 219$ ) and (b) many students attrited before completing our focal problems ( $n = 593$ ; see From Here to There! section for more detail). Of these 593 students, 449 had completed the algebra knowledge scale. We note that these 449 attrited students had a somewhat lower pretest score ( $M = 4.21$ ,  $SD = 2.6$ ) than the 837 who completed through our focal problems ( $M = 4.87$ ,  $SD = 2.6$ ); we return to this in the Discussion.

Because we used a convenience sample from a larger efficacy study, we conducted a post-hoc power analysis based on results from our earlier pilot study ( $n = 217$ ; Bye et al., 2022) and the full new sample size of 837. Specifically, we carried out the RQ1-2 data analyses described below using data from the pilot study conducted with 217 students in the same school district using FH2T a year prior to the current study sample; we extracted the estimated standard errors for parameters of interest, and multiplied them by the square root of 217/837 to account for the size of the new sample. We used these anticipated standard errors for power calculations. For RQ1, we estimated > 99% power to detect an effect size similar to what was observed in the pilot data, at  $\alpha = 0.05$ . The minimum detectable effect size at 80% power is an odds ratio of 1.39 (or 0.72), which is below the "weak association" benchmark of Chen et al. (2010) and approximately one-fifth of the effect size observed in the pilot data (Bye et al., 2022). For RQ2, we estimated 99% power to detect an effect size similar to the largest effect observed in the pilot data (an effect related to the PMET measure), at  $\alpha = 0.05$ . The minimum detectable effect size at 80% power is an odds ratio of 1.17 (or 0.85), about two-thirds of the largest effect

observed in the pilot data. The analyses for RQ3 are analogous to those for RQ1-2, so the power and minimal detectable effect sizes are the same.

Of the 837 students, 54% were identified by the district as boys, 46% as girls. Regarding race/ethnicity, 52% identified as White, 23% as Asian, 18% as Hispanic/Latinx, 4% as Black, 3% identified as other ethnicities, and 1% had missing race/ethnicity information. In terms of other demographic information, 11% of the participants were classified as English for Speakers of Other Languages (ESOL), 17% identified as gifted students, and 11% were classified as Individualized Education Program (IEP) students. As this study was conducted during the COVID-19 pandemic, students had the choice between attending school fully in person or virtually. Regarding classroom format, 67% were initially enrolled in in-person classrooms, 32% were in virtual classrooms, and 1% had missing information.

## Procedure

In the larger randomized controlled study, all students completed a 40-minute pretest on mathematics knowledge including a math anxiety assessment, up to nine 30-minute intervention sessions solving problems in an online mathematics game, and a 40-minute posttest as a part of their mathematics instruction across the school year (For more information on the larger study, see [Decker-Woodrow et al., 2023](#)). All study assignments were administered online either in math classrooms during instructional periods (for students attending school) or at home (for virtual students). In each session, teachers assigned a link and students worked individually on the assignment using a device. For students who were randomly assigned to FH2T (see below), they played through the game for 30 minutes at their own pace, up to nine times during the school year. As students played, the game automatically logged all their user actions. In each subsequent session, students continued the game starting on the problem where they left off.

## From Here to There!

From Here to There! (FH2T; [Ottmar et al., 2015](#)) is a research-backed interactive math game that allows students to dynamically manipulate notation and transform expressions to equivalent expressions using a number of gesture-actions. FH2T consists of 14 levels with 18 problems in each. Each level represents a core mathematical topic that increases in complexity as students play (e.g., addition and subtraction, order of operations, distribution and factoring, solving basic equations).

On each problem, students are presented with two expressions—a start state, which is transformable, and a goal state, which is perceptually different but mathematically equivalent to the start state ([Figure 1](#)). The objective of each problem is to transform the starting expression to the goal state. A sample problem with a series of actions is presented in [Figure 1](#). In this example, students are asked to transform the starting state ( $8 + 4 \times 4 \times 8$ ) to the goal state (136; [Figure 1a](#)). Students can interact with the notations, and the system provides immediate feedback on student actions. If students attempt invalid actions, such as tapping the addition sign to combine 8 and 4 in  $8 + 4 \times 4 \times 8$  ([Figure 1b](#)), the expression shakes ([1c](#)) and remains as  $8 + 4 \times 4 \times 8$  ([1d](#)), so students can try another action. If students attempt a valid action, such as tapping the multiplication sign in the middle ([Figure 1e](#)), the expression automatically undergoes a fluid transformation (e.g., [1e-1g](#)). Students can take any series of mathematically valid steps that transform the initial expression to the goal. However, the step count (in the bottom right next to the goal; [Figure 1](#)) increases as students perform valid actions, and turns red when students' steps exceed the minimum number required to complete the problem. After completing each problem, students receive either one, two, or three clovers. Students receive more clovers as their strategies are more efficient. As students play, all valid and invalid actions and time stamps are recorded in log data, providing a means to recreate and analyze students' math steps, strategies, and errors for each problem.

**Figure 1**

*A Sample Problem and Students' Actions in From Here to There!*



### Specific FH2T Problems Used in the Analysis

Here, we focused our analyses on six focal problems from World 3: *Order of Operations: Addition and Multiplication* (Table 1). These problems were chosen because they were carefully constructed to vary the position of addition and multiplication operations, i.e., three problems have only one addition operator (in each of the three locations) and three have only one multiplication operator (in each of the three locations).

**Table 1**

*The Six Target Problems and Whether Their Leftmost Operation Is Valid*

Problem Number	Start State (S)	Goal State (G)	Valid Leftmost Operation
38	$3 \times 7 + 7 + 3$	31	Yes
39	$8 + 4 \times 4 \times 8$	136	No
40	$2 \times 3 + 3 \times 2$	12	Yes
41	$5 + 2 + 2 \times 5$	17	Yes
42	$4 + 1 \times 1 + 4$	9	No
43	$1 \times 6 \times 6 + 1$	37	Yes

*Note.* Problem numbers indicate the ordinal position of the problems within the game. Among these six problems, students started with Problem 38 and ended with Problem 43. For Problems 39 and 42, students could not perform the leftmost operation first.

In all six problems, students were instructed to simplify the start state ([S]) into a numerical value (i.e., goal state [G]). The problems varied in whether students could solve them from left to right. For Problems 38 and 43 (Table 1), students could perform all operations from left to right and each step would be mathematically valid. For Problems 40 and 41, performing the operations from left to right would result in a valid first action and an invalid second action. For Problems 39 and 42, attempting the first action on the left would be mathematically invalid.



All students received the problems in the same order (indicated by problem number). We then analyzed the raw log data collected for those specific problems as students used mouse- or keyboard-actions to transform and simplify expressions.

## Measures

### Student Action Log Data

FH2T log data recorded all students' actions and problem-solving processes on each problem. Specifically, each time a student interacted with an element in an expression by clicking with a mouse or pressing on a touchscreen, the server logged: a) which element the student clicked/pressed on originally (e.g., the leftmost '8' or the '+' operator in Problem 39), b) whether the interaction was a 'tap' (e.g., clicking the '+' and releasing in the same location) or a 'drag' (e.g., clicking the '8', dragging it onto the '4', and then releasing), c) the timestamp of the action, d) the type of mathematical action (e.g., addition, multiplication, commutation, or a mistake, see below), and e) the resulting expression if a valid action (transformation) was made (e.g., "8 + 16 × 8" if the two 4s were multiplied together in Problem 39). Because of this detailed log structure, we were able to examine step-by-step mathematical processes and behaviors of students, including how students carried out valid transformations and what kinds of mistakes they made; these latter interactions were classified by whether they made precedence errors (see below).

Because students progressed through problems at their own pace across multiple sessions in class, in rare cases they first saw a problem and only finished it on a later date. There were only 10 such cases (out of 5,022 possible; i.e., 6 problems for each of 837 students) in which a student interacted with a problem before returning and finishing it on a later date; these 10 cases (from 10 different students) were thus treated as missing data in the present analyses. If instead a student merely saw a problem on one day and returned to interact with and finish it another day, then the data from the latter day was used. In all other cases, students finished the problem on their first viewing. Many students returned to problems days or months later to complete them a second or third time; in such cases, only the first attempt is used in the present analysis.

### Precedence Errors

For RQs 1-2, we examined all student interactions with the start state of each problem's expression (Table 1), prior to any valid action (i.e., transformation; Figure 1e to 1g). For RQ3, we examined their interactions after one valid interaction but before the second valid action. The raw log data itself recorded as a 'mistake' any action that did not transform the expression to a new state (i.e., is not a valid action). We characterized a 'precedence error' as an instance of a mistake in the log data where a student invalidly attempts an addition operation, e.g., by tapping on a '+' sign or dragging one addend onto the other, when this conflicts with precedence. For example, in Problem 39, attempting to add the first 8 and 4 together by tapping the '+' (Figure 1b) or dragging one number onto the other would generate a shake animation (1c); the expression would remain in the current state (1d) because multiplication (1e-1i) takes precedence before addition (1j-1l). Thus, we processed all students' log data for all problems, identifying occurrences where logged mistakes indicate an attempt at addition (by tapping or dragging) when it violates precedence (i.e., an adjacent operation involves the higher-order multiplication operation). Where these 'precedence errors' occurred prior to the students' first valid transformation, they were included as the outcome variables in analyses for RQs 1-2 and coded as a '1' (a '0' indicated they never made a precedence error prior to the first transformation); where they occurred after the students' first valid transformation but before their second, they were included in the analyses for RQ3 (again coded as 1 or 0). Since it was possible for students to 'reset' the expression back to the starting state and 'undo' their first valid action, we did not consider precedence errors that occurred after a reset or undo for RQ3. In 77 unique cases (1.5% of the 5,012 total observations), a student hit 'reset' after their first transformation but before their second. Since this does not provide opportunity to observe a precedence error (or its absence) at this stage, these 77 cases are removed from RQ3 analyses, leaving 4,941 valid observations for RQ3.

### Perceptual Math Equivalence Task (PMET)

We developed the PMET, a timed computerized task individually administered to middle school students, in order to measure students' ability to perceive math equivalence in algebraic expressions. As explained above, this is an important task that students routinely struggle with (NAEP, 2022). Grounded in prior research (Kellman et al., 2010; Marghetis et al., 2016), the PMET consists of two parts with eight items each. As adapted from Marghetis and colleagues' math equivalence task (2016), for each item in PMET Part 1, students saw two expressions and determined whether they were equivalent or not equivalent (Figure 2a). As adapted from Kellman and colleagues' training of perceptual learning modules (2010), for each item in Part 2, students saw a target expression and six options, and selected the option that was equivalent (four items; Figure 2b) or not equivalent to the target (four items). We intended Part 2 to be more difficult than Part 1, as it involved comparing among seven expressions (instead of just two); this design allowed us to capture a wider range of students' abilities. We also note that Part 2 bears similarity to the recent NAEP items on which students struggled (see Intro). The target population for PMET administration is students in middle and high school mathematics.

**Figure 2**

Sample Items in the First (a) and Second (b) Part of the PMET Task

<p><b>a)</b></p> <p style="text-align: center;">Are these two expressions equivalent?</p> <table style="width: 100%; text-align: center; border-collapse: collapse;"> <tr> <td style="border: 1px solid black; padding: 5px;"><math>2 \cdot (3y - 6)</math></td> <td style="border: 1px solid black; padding: 5px;"><math>-2 \cdot 6 + 2 \cdot 3y</math></td> </tr> </table> <table style="width: 100%; text-align: center; margin-top: 10px;"> <tr> <td style="border: 1px solid black; padding: 5px;">Equivalent</td> <td style="border: 1px solid black; padding: 5px;">Not Equivalent</td> </tr> </table>	$2 \cdot (3y - 6)$	$-2 \cdot 6 + 2 \cdot 3y$	Equivalent	Not Equivalent	<p><b>b)</b></p> <p style="text-align: center;">Which expression is equivalent to the following?</p> <p style="text-align: center; font-size: 1.2em;"><math>y - (-8)</math></p> <table style="width: 100%; text-align: center; border-collapse: collapse;"> <tr> <td style="border: 1px solid black; padding: 5px;"><math>-y - 8</math></td> <td style="border: 1px solid black; padding: 5px;"><math>-y + 8</math></td> <td style="border: 1px solid black; padding: 5px;"><math>y - 8</math></td> </tr> <tr> <td style="border: 1px solid black; padding: 5px;"><math>8 + y</math></td> <td style="border: 1px solid black; padding: 5px;"><math>8 - y</math></td> <td style="border: 1px solid black; padding: 5px;"><math>-8 + y</math></td> </tr> </table>	$-y - 8$	$-y + 8$	$y - 8$	$8 + y$	$8 - y$	$-8 + y$
$2 \cdot (3y - 6)$	$-2 \cdot 6 + 2 \cdot 3y$										
Equivalent	Not Equivalent										
$-y - 8$	$-y + 8$	$y - 8$									
$8 + y$	$8 - y$	$-8 + y$									

It is important to note that students were not asked to solve equations in PMET; instead, they were to attend to the algebraic structure of, and shift attention between, various expressions or equations, in order to identify expressions or equations that are mathematically equivalent and can be connected through valid algebraic transformations. Students' accuracy (1 = correct; 0 = incorrect) and response time (RT) on each item were recorded. KR-20 (Kuder-Richardson Formula 20) for all 16 items was 0.68. We used the total accuracy scores (0 to 16) and total time (i.e., time taken to finish all of the problems, in seconds) of PMET at pretest as two predictors for RQ2 and RQ3 in the analyses. A higher accuracy score represents better ability in perceiving math equivalence in algebraic expressions, whereas a higher RT represents slower and potentially less fluent ability in perceiving math equivalence. By including both accuracy and RT in the models, we are able to tease apart these two aspects of a speeded task.

### Algebra Knowledge Assessment

The algebra assessment consisted of 10 multiple-choice items selected from a previously validated measure (Star et al., 2015; Cronbach's  $\alpha = .89$ ). Capturing a more comprehensive understanding of algebra compared to PMET, this algebra knowledge assessment tapped on students' conceptual knowledge (4 items; definition of the equal sign), procedural knowledge (3 items; equation solving), and procedural flexibility (3 items; identify efficient strategies) of algebra

concepts. A sample item was “solve the equation for  $y$ :  $5(y - 2) = 3(y - 2) + 4$ .” Each item was scored as correct (1) or incorrect (0), and the reliability of these items were  $KR-20 = .74$  in the full sample ( $n = 1,850$ ). Each student’s total sum score (0-10) was included as a covariate for RQ2 and RQ3 in the analyses.

### Math Anxiety

Students’ math anxiety was measured using 9 items selected from the Math Anxiety Scale for Young Children-Revised ( $\alpha = 0.87$ ; Ganley & McGraw, 2016). A sample item was “math gives me a stomachache”. Students rated how well each statement described their feeling on a four-point scale (no = 0; not really = 1; kind of = 2; yes = 3). The reliability of the scale from the full sample was  $\alpha = 0.83$ . Students’ average scores on the math anxiety scale at pretest were included as a covariate for RQ2 and RQ3 in the analyses.

### Analytic Approach

First, we conducted descriptive and correlational analyses to examine the distribution of and relations between variables at the student level. These preregistered analyses (detailed below) involved a series of mixed effects binary logistic regressions to predict the occurrence of a precedence error (a binary outcome defined below) for each student on each of the six target problems. All models included crossed random effects for students and for problem numbers (Baayen et al., 2008), along with fixed effects corresponding to the research question.

As a null model for the RQs, we fit an unconditional model to estimate the proportion of variance in precedence errors due to each unique student and problem, modeling precedence errors as a binary outcome with crossed random effects for student and problem (Baayen et al., 2008). For RQ1, we predicted the probability of making precedence errors prior to a valid action, as a function of a problem-level predictor (i.e., whether the leftmost operation is valid). For RQ2, we first added the algebra knowledge assessment score and math anxiety score as student-level predictors to the model. We then subsequently added PMET accuracy scores and response time to the models as measures of students’ perception of equivalent expressions. For RQ3, we repeated analogous models to RQs 1-2 for precedence errors occurring between students’ first and second valid actions.

Following helpful feedback from two reviewers, we deviated from our preregistration (Supplementary Materials) to A) measure students’ aggregate performance on the PMET rather than each section separately, in line with best practices, and B) also include PMET response time to account for time-on-task, as distinct from accuracy. Given that RT was found to be positively skewed (as is typical with RT data), we used a log transform on RT. All other analysis details follow the preregistration, unless noted otherwise. All analyses were performed in R (R Core Team, 2020) with the tidyverse (Wickham et al., 2019) and the lme4 packages (Bates et al., 2015), with model fits computed with the performance package (Lüdtke et al., 2021). All data and code for these analyses is available on OSF (see Bye et al., 2024S).

## Results

Table 2 presents means, standard deviations, and correlations (with confidence intervals) of the variables included in the analyses. As shown in Table 2, the proportion of precedence errors prior to the first valid action had significant negative correlations with algebra knowledge ( $r = -.17, p < .001$ ), math anxiety ( $r = .08, p = .017$ ), PMET accuracy ( $r = -.22, p < .001$ ), and PMET Log RT ( $r = -.14, p < .001$ ). The proportion of precedence errors after students’ first valid action but prior to their second action also had significant negative correlations with algebra knowledge ( $r = -.17, p < .001$ ), PMET accuracy ( $r = -.16, p < .001$ ), and PMET Log RT ( $r = -.10, p = .003$ ). However, there was no relation between these precedence errors and math anxiety ( $r = .04, p = .25$ ). In line with a speed-accuracy tradeoff, participants with longer PMET RTs were more likely to answer PMET items correctly ( $r = .39, p < .001$ ).

Table 2

Means, Standard Deviations, and Pearson's Correlations [With Confidence Interval] for the Variables of Interest

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5
1. Proportion of Precedence Errors (1st)	0.11	0.15					
2. Proportion of Precedence Errors (2nd)	0.07	0.10	.23*** [.17, .30]				
3. Algebra Knowledge	4.87	2.62	-.17*** [-.24, -.10]	-.17*** [-.24, -.10]			
4. Math Anxiety	1.52	0.65	.08* [.01, .15]	.04 [-.02, .11]	-.23*** [-.29, -.16]		
5. PMET accuracy	7.56	3.00	-.22*** [-.28, -.15]	-.16*** [-.23, -.09]	.65*** [.61, .69]	-.20*** [-.26, -.13]	
6. PMET log RT	5.74	0.80	-.14*** [-.20, -.07]	-.10** [-.17, -.04]	.38*** [.32, .44]	-.07* [-.14, .00]	.39*** [.34, .45]

*Note.* The first two variables represent the proportion of problems on which participants made precedence errors (ranging from 0 to 1); the proportion is used instead of the number of precedence errors because there were different amounts of missing data between the two time points (see *Measures* above). The first variable is the proportion of precedence errors prior to the first valid action; the second variable is the proportion after the first valid action but before the second. Variables 3-6 are taken from the pretest (prior to using FH2T). Variable 6 is log-transformed RT (sec).

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

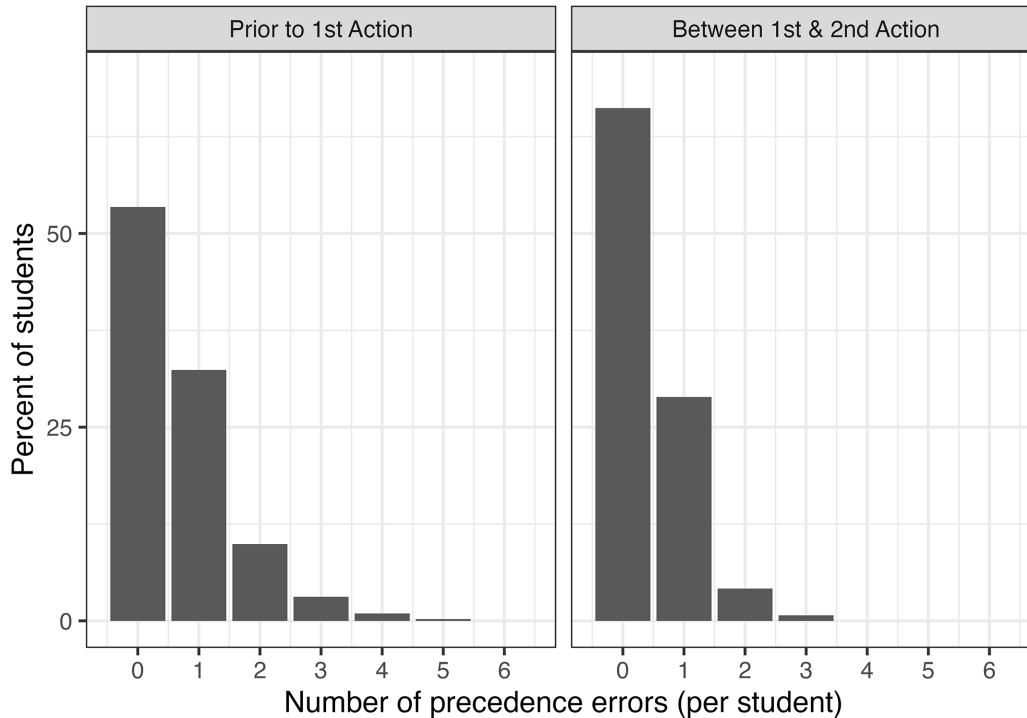
We also examined how many students made precedence errors prior to their first valid action (Figure 3, left), finding that just over half of the students made no such errors at all (53%;  $n = 447$  of 837) and 47% of students ( $n = 390$ ) made precedence errors prior to their first valid action. In total, about a third (32%;  $n = 271$ ) made a precedence error on only one problem and about a tenth ( $n = 83$ ) of the students made precedence errors on two. The remaining 36 students made errors on 3-5 problems, but none made errors on all 6. Overall, students averaged 0.66 precedence errors across all 6 problems, prior to their first valid action; this translated to an average proportion of 0.11 precedence errors *per* problem ( $SD = .15$ ; see Table 2).

For the precedence errors between students' first and second valid actions (Figure 3, right), most students made no such precedence errors (66%;  $n = 554$  of 837). In total, over a quarter of students (29%;  $n = 242$ ) made such a precedence error on only one problem, and about 4% ( $n = 35$ ) of the students made them on two problems. Only 6 students made these errors on three problems, and none made them on more than three problems. Overall, students averaged a total of 0.4 precedence errors across all 6 problems between their first and second valid action, i.e., they averaged a rate of 0.07 precedence errors per problem ( $SD = .10$ ; see Table 2).

For RQ1 and RQ2, we focused on precedence errors prior to the first valid action. As a baseline for the subsequent hierarchical logistic regressions, we fit an unconditional null model predicting precedence errors (a binary outcome). There were six problems completed by all 837 students, for a total of 5,012 observations (10 additional observations were treated as missing, see above). Since each student completed each problem, we modeled both student and problem as crossed random effects (Baayen et al., 2008). We first ran models with each random effect on its own to quantify their individual contributions, revealing that the conditional  $R^2$  for the random effect of student was 15.9% and the random effect of problem was 10.2%, indicating both random factors matter. We then ran the full null model with both crossed random effects, revealing a conditional  $R^2$  of 23.4%, suggesting a substantial amount of variance explained in precedence errors,  $p < .001$  (see Model 1.0 in Table 3); therefore, in all following analyses, we included both random effects.

**Figure 3**

Frequency of Precedence Errors by Student (a) Prior to First Valid Action and (b) Between First and Second Valid Actions



Note. Barplot representing the percent of students in the sample who made precedence errors on 0 to all 6 problems, prior to the first valid transformation (left facet) or between their first and second transformations (right facet).

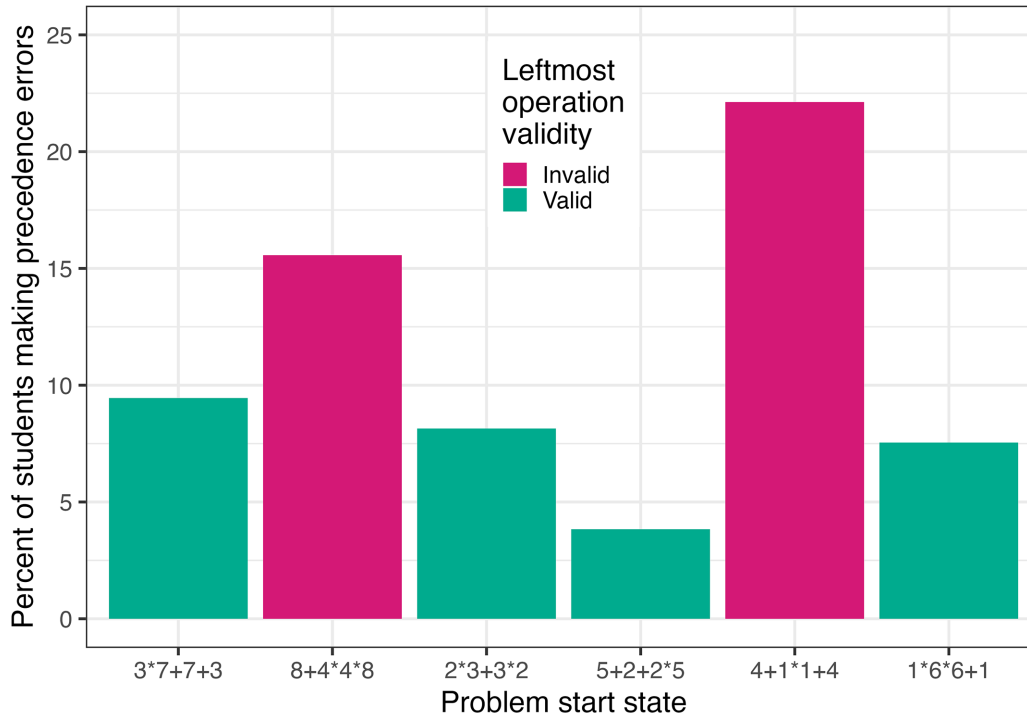
### RQ1: Effects of Problem Structure on Students' Precedence Errors

Our first research question concerned whether the problem structure (i.e., whether performing the leftmost operation is a valid or invalid first action) affects the propensity for students to make precedence errors. As predicted, when performing the leftmost operation was an invalid first action, the odds of students making a precedence error were 228% higher than when the leftmost operation was higher in precedence,  $OR = 3.28$ , 95% CI [2.12, 5.07],  $z = 5.35$ ,  $p < .001$  (see Model 1.1 in Table 3; compare the Invalid problems to Valid in Figure 4).

In a separate analysis of just the two problems in which performing the leftmost operation was an invalid first action (39 and 42; see pink bars in Figure 4), we examined whether students who committed precedence errors on one problem (i.e., Problem 39) were more likely than their peers to commit an error on the other (i.e., Problem 42). Specifically, we found that the odds of students to make a precedence error on Problem 42 were 104% higher for those who had previously made a precedence error on Problem 39 than those who had not,  $OR = 2.04$ , 95% CI [1.35, 3.06],  $z = 3.44$ ,  $p < .001$ . In total, 227 students (27.1%) made a precedence error on exactly one of the problems and 44 students (5.3%) made a precedence error on *both* problems.

Figure 4

Percent of Students Making Precedence Errors by Problem, Prior to First Valid Action



Note. Problems are listed left to right in the order in which students saw them in the FH2T game. The two problems where the leftmost operation is initially invalid are indicated in pink, while the four problems with a valid leftmost operation are indicated in green.

## RQ2: Effects of Students' Pretest Measures on Their Precedence Errors

Our second research question concerned students' pretest measures (algebra knowledge, math anxiety, PMET accuracy and response time) and to what extent they predicted students' precedence errors. All four pretest measures were mean-centered.

Given our primary interest in PMET and the well-established literature on both algebra knowledge and math anxiety measures, we first add algebra knowledge and math anxiety as covariates to account for their influences (see Model 1.2 in Table 3). Each algebra question answered correctly was associated with a 10% decrease in the odds of students to make a precedence error,  $OR = 0.90$ , 95% CI [0.86, 0.94],  $z = -4.70$ ,  $p < .001$ . The math anxiety pretest did not explain significant variance,  $z = 1.34$ ,  $p = .18$ .

We then added the PMET accuracy scores and response time to the model to see whether they explain variance over and above the covariates (Model 1.3). We found that each additional PMET item answered correctly was associated with a 10% decrease in the odds of students to make a precedence error,  $OR = 0.90$ , 95% CI [0.86, 0.95],  $z = -4.20$ ,  $p < .001$ . PMET log RT was not significantly related to the odds of students to make a precedence error,  $z = -1.29$ ,  $p = .20$ .

Importantly, adding PMET made the algebra knowledge assessment no longer significant in the model (compare Model 1.2 to 1.3). Due to this change in significance for other student-level predictors, we fit a post-hoc exploratory model with PMET accuracy as the only student-level fixed effect (in addition to problem-level leftmost operation validity and the random effects of student and problem). This reduced model explained similar variance in precedence errors as Model 1.3 even without the other three predictors, with a marginal  $R^2$  of .107, a conditional  $R^2$  of .243, and log-likelihood of -1624.4. Thus, among the pretest measures, students' PMET accuracy scores clearly explained the most variance in their precedence errors.

**Table 3***Model Comparison Table for the Hierarchical Logistic Regression Predicting Precedence Errors Prior to the First Valid Action*

Effects	Model 1.0	Model 1.1	Model 1.2	Model 1.3
<b>Fixed effects</b>				
(Intercept)	0.09*** [0.05, 0.14]	0.06*** [0.04, 0.08]	0.06*** [0.04, 0.08]	0.06*** [0.04, 0.08]
Left operation invalid	–	3.28*** [2.12, 5.07]	3.29*** [2.09, 5.18]	3.28*** [2.05, 5.25]
Algebra knowledge	–	–	0.90*** [0.86, 0.94]	0.98 [0.93, 1.03]
Math anxiety	–	–	1.13 [0.95, 1.34]	1.10 [0.93, 1.30]
PMET accuracy	–	–	–	0.90*** [0.86, 0.95]
PMET log RT	–	–	–	0.91 [0.79, 1.05]
<b>Random effects</b>				
SD (Stu.)	0.78	0.81	0.76	0.72
SD (Prob.)	0.62	0.23	0.24	0.25
<b>Model fits</b>				
Log-likelihood	-1652.1	-1647.1	-1632.9	-1623.4
AIC	3310.2	3302.1	3277.8	3262.9
Marg. $R^2$	–	.073	.094	.104
Cond. $R^2$	.234	.237	.240	.235

Note. Fixed effects report odds ratio [with 95% CI]. Random effects report standard deviation of intercept.

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

### RQ3: Precedence Errors Between the First and Second Valid Actions

We next applied parallel analyses to the precedence errors generated after students performed one valid action (e.g., multiplying the first 2 and 3 in  $2 \times 3 + 3 \times 2$  to yield  $6 + 3 \times 2$ ). We fit an unconditional null model predicting precedence errors between the first and second valid actions, with crossed random effects for each student and for each problem. As with RQs 1-2, there were six problems completed by 837 students. After removing the 10 missing observations from RQ1-2 and the additional 77 cases where a student hit ‘reset’ after the first transformation (see Measures), this left a total of 4,935 observations for the RQ3 analysis. This time, the random effect of problem accounted for 10.5% of the variance in precedence errors between the first and second valid actions (see Model 2.0 in Table 4). However, unlike for precedence errors prior to the first transformation (RQs 1 and 2), for precedence errors after their first transformation there was no apparent random effect of student (this model suffered from singularity); most likely, this is due to the high variability in the problem state across students within a given problem, overriding random effects of students (e.g., many students had transformed  $2 \times 3 + 3 \times 2$  into  $6 + 3 \times 2$ , while many others had transformed the same initial state into  $2 \times 3 + 6$ ). Thus, we proceeded to only model the random effect of problem for RQ3 to avoid singularity issues.

We next assessed whether the current structure of the problem after the first transformation affects students’ propensity to make precedence errors. Unlike in RQ2 where the six problems had only one initial state each, here all problem types could appear with a valid leftmost operation (e.g.,  $2 \times 3 + 6$ ) or not (e.g.,  $6 + 3 \times 2$ ). As predicted, when performing the leftmost operation was an invalid action, the odds of students making a precedence error were 428% higher than when it was valid,  $OR = 5.24$ , 95% CI [3.23, 8.50],  $z = 6.7$ ,  $p < .001$  (see Model 2.1 in Table 4).

**Table 4**

Model Comparison Table for the Hierarchical Logistic Regression Predicting Precedence Errors After the First Valid Action but Before the Second

Effects	Model 2.0	Model 2.1	Model 2.2	Model 2.3
<b>Fixed effects</b>				
(Intercept)	0.06*** [0.04, 0.10]	0.03*** [0.02, 0.06]	0.03*** [0.02, 0.06]	0.03*** [0.02, 0.06]
Left operation invalid	–	5.24*** [3.23, 8.50]	5.20*** [3.22, 8.41]	5.16*** [3.20, 8.34]
Algebra knowledge	–	–	0.89*** [0.85, 0.93]	0.93* [0.87, 0.98]
Math anxiety	–	–	1.02 [0.85, 1.22]	1.01 [0.84, 1.20]
PMET accuracy	–	–	–	0.95 [0.90, 1.00]
PMET log RT	–	–	–	0.94 [0.81, 1.10]
<b>Random effects</b>				
SD (Prob.)	0.62	0.71	0.71	0.71
<b>Model fits</b>				
Log-likelihood	-1169.4	-1142.8	-1129.7	-1127.4
AIC	2342.8	2291.6	2269.4	2268.7
Marg. $R^2$	–	.136	.152	.156
Cond. $R^2$	.105	.252	.266	.269

Note. Fixed effects report odds ratio [with 95% CI]. Random effects report standard deviation of intercept.

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

We next examined to what extent students' pretest measures predicted these precedence errors. All four pretest measures were again mean-centered. As above, algebra knowledge and math anxiety were entered first as covariates (see Model 2.2 in Table 4). Each additional algebra question answered correctly was associated with an 11% decrease in the odds of students making a precedence error,  $OR = 0.89$ , 95% CI [0.85, 0.93],  $z = -4.9$ ,  $p < .001$ . The math anxiety pretest did not explain significant variance over and above the previous model,  $z = 0.17$ ,  $p = .86$ .

We then added the PMET accuracy scores and response time to the model to see whether they explain variance over and above the covariates (Model 2.3). Unlike in RQ2, PMET accuracy was not significant,  $OR = 0.95$ , 95% CI [0.90, 1.00],  $z = -1.84$ ,  $p = .066$ . Again, PMET log RT was not significantly related to the odds of students to make a precedence error ( $p = .45$ ). Here, while adding PMET scores did decrease the effect of algebra knowledge, it did so to a lesser degree than RQ2, with algebra knowledge remaining significant in RQ3.

## Discussion

Many students struggle to reconcile precedence (i.e., the order of operations) with a strong tendency for left-to-right computation. In this study, we utilized data logged as students solved problems in an online, interactive math game to examine individual students' propensity to make precedence errors across problems varying in syntactic structure. In line with prior research, we found that students varied in their susceptibility to precedence errors, and the frequency



of these errors varied with individual problem structures (when left-to-right problem-solving approach deviates from precedence; RQ1). We also found that precedence errors on students' initial actions (i.e., prior to any valid operation) were significantly predicted by students' scores on a task measuring perception of equivalence, to a greater extent than a typically used measure of procedural and conceptual algebra knowledge (RQ2). These results replicated findings from our prior work using a different smaller sample (Bye et al., 2022). Finally, we found that these same effects partially replicated after students have performed their first valid action (e.g., transformed  $2 \times 3 + 3 \times 2$  into  $6 + 3 \times 2$ ; RQ3). Although the problem states were much more variable after the first transformation, the propensity to make precedence errors still significantly varied across problem structures (i.e., whether the leftmost operation was valid). However, students' perception of equivalence was no longer a significant predictor of precedence errors after their first valid action. Below we discuss each of these findings and their implications for research and practice.

## Syntactic Structure of Expressions Influence Probability of Precedence Errors

The most robust and largest predictor of students' precedence errors was whether the leftmost operation was initially invalid (e.g.,  $8 + 4 \times 4 \times 8$ ) vs. valid (e.g.,  $3 \times 7 + 7 + 3$ ); this predictor remained significant and substantial across all models. Although students could make precedence errors on any of the six problems, they were considerably more likely to do so when the leftmost operation was invalid (Figure 4). This finding is consistent with prior research (Banerjee & Subramaniam, 2005; Blando et al., 1989; Gunnarsson et al., 2016; Norton & Cooper, 2001) and our own pilot with a smaller sample (Bye et al., 2022). As McNeil (2014) has argued, students' experience with arithmetic is narrow, leading to their inflexibility in problem solving and resistance to change their strategies. Aligned with this view, we found that for some students, entrenchment in solving problems from left to right persisted even when doing so would violate the order of precedence. Among our sample of 837 middle school students, 15%-22% of the students continued to solve problems from left to right and committed a precedence error on the problems where the leftmost operation was lower in precedence (see Figure 4). And while making a precedence error on problem #39 predicted making one on problem #42 (see RQ1), only 5.3% of students ( $n = 44$ ) made a precedence error on both problems. These findings suggested that some middle-school students struggled to break away from the left-to-right habit, which is needed to develop flexibility in algebraic problem solving, but most students do not always fall for an invalid leftmost operation. As individuals varied significantly on their propensity to make precedence errors, we further investigated predictors of precedence errors.

## Individual Predictors of Probability in Making Precedence Errors

Although 53% of students did not make any precedence errors prior to their first valid action, 32% did make an error on one problem, 9.9% made errors on two problems, and 15% made errors on three to five problems before performing a valid action. The notable differences in individuals' tendency to make precedence errors was further supported by the multilevel model in which 15.9% of the variance was attributable to students. So what factors accounted for these individual differences in making precedence errors? Grounded in prior literature (e.g., Foley et al., 2017; Hong et al., 2023; Kellman & Massey, 2013; Spielhagen, 2006), we examined students' perception of math equivalence, conceptual and procedural knowledge of algebra, and math anxiety as potential predictors of the likelihood that people would make precedence errors. Students' perception of equivalence, as measured by PMET, significantly predicted their probability in making precedence errors. Specifically, each additional item answered correctly on PMET was associated with a 10% decrease in the odds of students making a precedence error prior to their first valid action. While students' algebra knowledge was also predictive of precedence errors, errors were more strongly predicted by PMET versus algebra knowledge prior to the first valid action (but not the second). Students' math anxiety was not predictive of precedence errors, nor was their time spent on the PMET.

Prior research has demonstrated that the role of perceptual grouping of symbols within expressions affects children's and adults' performance on order of precedence problems (Braithwaite et al., 2016; Harrison et al., 2020; Landy & Goldstone, 2010). Our findings, however, are the first to show that, in addition to the perceptual features of the expressions, students' abilities to perceive equivalent expressions are also associated with their likelihood of adhering to the order of precedence. While this study does not allow us to definitively identify the mechanisms behind these

relations, one potential interpretation is that the PMET measure captures the ability to quickly and accurately perceive algebraic structures among various transformations of equivalent expressions, and this ability may support students' attention to algebraic structures of expressions, resulting in adherence to the order of operations and overall equation solving performance. Furthermore, it is plausible that this perception comes from repeated problem-solving practice in the first place (i.e., it could be a bidirectional relationship; Kellman et al., 2010).

Alternatively, crucial skills such as executive function may account for the association between PMET and the probability of making precedence errors in FH2T. On one hand, PMET requires students to flexibly shift attention between several expressions or equations and to recognize algebraic transformations of these expressions or equations (also see Hong et al., 2023 on math flexibility). On the other hand, avoiding precedence errors requires students to attend to the entire expression and exercise inhibition, especially when performing operations from left to right is an invalid strategy. A large body of work has investigated the relations between executive function skills and mathematics and found mixed results (e.g., Bull & Lee, 2014; Chan & Scalise, 2022; Cragg & Gilmore, 2014; Lee & Lee, 2019; Mazzocco et al., 2017; Medrano & Prather, 2023). Although we cannot tease apart the influences of perception of equivalence and executive function or rule out the potential involvement of executive function in our PMET, performance on items similar to those of PMET Part 1 have been found to associate with adults' perception of algebraic structures (Marghetis et al., 2016). Furthermore, perceptual training on the mapping between representations using items similar to those of PMET Part 2, have been found to be distinct from declarative or procedural knowledge of algebra and contribute to high school students' equation solving fluency (Kellman et al., 2010). Together, the prior research and current findings provide support for the influences of perception of equivalence on equation solving, particularly on students' first action.

Contrary to PMET, algebra knowledge was not a significant predictor of students' tendency to make precedence errors on their first action when all predictors were included in the model. One potential explanation for the lack of significant effect of algebra knowledge is its moderate to strong associations with PMET ( $r = .65$ ). It is also true that the algebra knowledge pre-assessment captures a broader set of abilities compared to PMET, as it includes items on solving for a variable and selecting efficient problem-solving strategies, in addition to identifying equivalent expressions. The wide range of conceptual, procedural, and flexibility items may not all be closely related to students' understanding of precedence and their tendency to make precedence errors, potentially contributing to the pattern of results we observed in the current study. The broader algebra knowledge scale was included as part of the larger RCT to also assess learning from the full intervention (Decker-Woodrow et al., 2023). We believe it is important to compare PMET to a more general algebra knowledge measure like this, given the clear link between understanding of precedence and performance in algebra (Linchevski & Livneh, 1999; Pillay et al., 1998; Spitzer & Moeller, 2022). Indeed, when looking at students' precedence errors between their first and second transformation, algebra knowledge remained a significant predictor while PMET did not. However, the leftmost operator validity was a descriptively stronger predictor here ( $OR = 5.16$  in the full model) than prior to the first transformation ( $OR = 3.28$ ). Given the wide variety of problem states encountered by students at this state (but not the initial state), and the fact that there was no apparent random effect of student in these models, it may be that students' individual differences play a smaller role than the problem state after a transformation. Future work should examine this nuanced relationship in more detail.

Interestingly, in all models, math anxiety never significantly predicted students' propensity of making precedence errors. In fact, its correlation with students' precedence errors was weak, and only significant for initial errors (Table 2). One possibility is that the gamified environment of FH2T encourages students to explore algebraic expressions, and potentially alters the ways in which learning contexts trigger students' math anxiety. Although speculative, this interpretation is supported by the finding that students' math anxiety was correlated with their hint usage (a measure of within-system interaction) in a traditional online homework platform but not in FH2T (Decker-Woodrow et al., 2023). In summary, the current study builds off work in perceptual learning (Kellman et al., 2010) and highlights the importance of students' perception of equivalence (as measured by PMET), providing exciting future directions to examine its potential role in mathematical learning and performance. Future studies should further investigate the relations among various predictors as well as the influences of context on the current findings to advance theories and research on students' algebraic problem solving.

## Implications for Research and Practice

The methods and findings from this study provide implications for both applied cognitive psychology research as well as instructional practice in mathematics education. First, this study provides a model for future applied research by analyzing rich log data that was collected within an interactive mathematics game embedded in a classroom context. Interactive technology interventions can capture a wide variety of students' mathematical problem solving processes and record behaviors at the level of their interactions with symbols in each expression. Furthermore, the predictive relation between PMET performance and precedence errors demonstrates the role of students' perception of equivalence in problem solving beyond their prior knowledge (in line with perceptual learning; Kellman et al., 2010; Marghetis et al., 2016). We acknowledge that our PMET measure is still in development and need to be further validated; however, the current findings motivate future research to include measures of perception of equivalence in order to further investigate its influences on mathematical problem solving and development. Together, this work demonstrates how traditionally lab-based tasks can be adapted and integrated into online math activities to gain more insights about student perception and cognition in authentic learning environments.

Second, the findings, along with prior literature, provide implications for teachers and content developers to consider variations in problem structures while designing mathematics activities for students. The findings suggest that when simplifying expressions that are similar in difficulty, students struggle more often when solving problems that cannot be solved through left-to-right calculations. This means that in practice, students should see a variety of problem structures and receive immediate feedback on why errors are occurring to help encourage more flexible problem-solving strategies. Further, using a variety of problem structures in both practice and assessments will likely provide more accurate representations of students' knowledge and flexibility, since these practice problems and assessments will reveal greater problem-level variability in students' performance.

## Limitations and Future Directions

The current study has several limitations. First, due to the pre-set structure of the game's levels, all students solved the target problems in the same order, which might have created an undetected ordering effect. Future work could build on our findings by ensuring randomization of problems to rule this possibility out. Second, the data used in this study were collected during the height of the COVID-19 pandemic, where students exhibited higher than usual levels of attrition, absence, remote learning, disengagement, and learning loss. Due to missing pretest data, we were unable to determine whether students who dropped out of the larger study were missing at random for all possible factors that were pertinent to our results; however, as noted in Participants, attrited students who had completed the algebra knowledge scale did score somewhat lower than students in our final sample, creating a limitation to our ability to generalize to all students. Nonetheless, since our primary outcome measure involves precedence errors, it is plausible that these effects may be even stronger among students with lower algebra knowledge. Third, our data came from a gamified educational technology rather than a more traditional mathematics assessment which might have impacted our ability to generalize our approach to all educational contexts. Finally, while the PMET is based directly off of prior research in perceptual learning (Kellman et al., 2010; Marghetis et al., 2016), more advanced psychometrics are needed before wider spread adoption of such a measure. Although the current study was limited in these ways, the findings nonetheless provided insights into the associations between problem structure, perception of equivalence, and students' equation solving performance, and advanced research that delineates these associations to inform intervention development. For example, what kinds of perceptual training (e.g., Kellman et al., 2010) might help students avoid the overapplication of left-to-right problem solving?

## Conclusions

Through examining students' precedence errors in a math game, we highlight the persisting left-to-right strategies among some middle school students, the influences of problem structures on those precedence errors, and the potential role of students' perception of equivalence in their equation-solving performance. These findings extend prior research by contributing novel insights into factors that relate to students' fluency with precedence, and inform future research and practice that aim to support students' problem-solving flexibility.

**Funding:** The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through an Efficacy and Replication Grant (R305A180401) and the National Science Foundation through a NSF CAREER Grant (142984) to Worcester Polytechnic Institute. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

**Acknowledgments:** We thank the participating teachers and students for their support with this study. We also thank the participating teachers and students, Erik Weitnauer and David Brokaw at Graspable Inc., and members of the Math Abstraction Play Learning and Embodiment Lab for their work on this project. We also thank Adam Sales, Tai Do, and Andy Zieffler for statistical advice.

**Competing Interests:** The authors have declared that no competing interests exist.

**Ethics Statement:** This work has been carried out in accordance with relevant ethical principles and standards where IRB approval was obtained at Worcester Polytechnic Institute. Authors outside of WPI signed data sharing agreements with WPI.

**Data Availability:** All data and code for this study are publicly available on OSF (see Bye et al., 2024S).

## Supplementary Materials

The Supplementary Materials contain the following items:

- Preregistration (Bye et al., 2023S)
- Research data, codebook, and code (Bye et al., 2024S)

### Index of Supplementary Materials

Bye, J. K., Chan, J. Y.-C., Closser, A. H., Lee, J.-E., Shaw, S. T., & Ottmar, E. R. (2023S). *Perceiving precedence* [Preregistration]. OSF Registries. <https://osf.io/jt27b>

Bye, J. K., Chan, J. Y.-C., Closser, A. H., Lee, J.-E., Shaw, S. T., & Ottmar, E. R. (2024S). *Perceiving precedence* [Research data, codebook, and code]. OSF. <https://osf.io/hsn6y>

## References

- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*(2), 224–237. <https://doi.org/10.1037/0096-3445.130.2.224>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Banerjee, R., & Subramaniam, K. (2005). Developing procedure and structure sense of arithmetic expressions. In H. L. Chick & J. L. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 121–128). University of Melbourne, Department of Science and Mathematics Education.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., Krivitsky, P. N., Tanaka, E., & Jagan, M. (2015). Package “lme4.” *Convergence*, *12*(1), Article 2.
- Blando, J. A., Kelly, A. E., Schneider, B. R., & Sleeman, D. (1989). Analyzing and modeling arithmetic errors. *Journal for Research in Mathematics Education*, *20*(3), 301–308. <https://doi.org/10.2307/749518>
- Braithwaite, D. W., Goldstone, R. L., van der Maas, H. L. J., & Landy, D. H. (2016). Non-formal mechanisms in mathematical cognitive development: The case of arithmetic. *Cognition*, *149*, 40–55. <https://doi.org/10.1016/j.cognition.2016.01.004>
- Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, *8*(1), 36–41. <https://doi.org/10.1111/cdep.12059>
- Bye, J. K., Lee, J.-E., Chan, J. Y.-C., Closser, A. H., Shaw, S. T., & Ottmar, E. R. (2022). *Perceiving precedence: Order of operations errors are predicted by perception of equivalent expressions* [Poster presentation]. Annual meeting of the American Educational Research Association (AERA), San Diego, CA, USA. <https://doi.org/10.3102/1884787https://doi.org/10.3102/1884787>

- Chan, J. Y.-C., Ottmar, E. R., Smith, H., & Closser, A. H. (2022). Variables versus numbers: Effects of symbols and algebraic knowledge on students' problem-solving strategies. *Contemporary Educational Psychology*, *71*, Article 102114. <https://doi.org/10.1016/j.cedpsych.2022.102114>
- Chan, J. Y.-C., & Scalise, N. R. (2022). Numeracy skills mediate the relation between executive function and mathematics achievement in early childhood. *Cognitive Development*, *62*, Article 101154. <https://doi.org/10.1016/j.cogdev.2022.101154>
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. [https://doi.org/10.1016/0010-0285\(73\)90004-2](https://doi.org/10.1016/0010-0285(73)90004-2)
- Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics – Simulation and Computation*, *39*(4), 860–864. <https://doi.org/10.1080/03610911003650383>
- Closser, A. H., Botelho, A., & Chan, J. Y.-C. (2024). Exploring the impact of symbol spacing and problem sequencing on arithmetic performance: An educational data mining approach. *Journal of Educational Data Mining*, *16*(1), 84–111.
- Closser, A. H., Chan, J. Y.-C., & Ottmar, E. R. (2023). Resisting the urge to calculate: The relation between inhibitory control and perceptual cues in arithmetic performance. *Quarterly Journal of Experimental Psychology*, *76*(12), 2690–2703. <https://doi.org/10.1177/17470218231156125>
- Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63–68. <https://doi.org/10.1016/j.tine.2013.12.001>
- Daker, R. J., Gattas, S. U., Necka, E. A., Green, A. E., & Lyons, I. M. (2023). Does anxiety explain why math-anxious people underperform in math? *npj Science of Learning*, *8*(1), Article 6. <https://doi.org/10.1038/s41539-023-00156-z>
- Decker-Woodrow, L. E., Mason, C. A., Lee, J.-E., Chan, J. Y.-C., Sales, A., Liu, A., & Tu, S. (2023). The impacts of three educational technologies on algebraic understanding in the context of COVID-19. *AERA Open*, *9*, Article 23328584231165919. <https://doi.org/10.1177/23328584231165919>
- Egorova, A., Ngo, V., Liu, A. S., Mahoney, M., Moy, J., & Ottmar, E. R. (2024). Mathematics presentation matters: How superfluous brackets and higher-order operator position in mathematics can impact arithmetic performance. *Mind, Brain and Education*, *18*, 258–269. <https://doi.org/10.1111/mbe.12421>
- Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The math anxiety-performance link: A global phenomenon. *Current Directions in Psychological Science*, *26*(1), 52–58. <https://doi.org/10.1177/0963721416672463>
- Ganley, C. M., & McGraw, A. L. (2016). The development and validation of a revised version of the math anxiety scale for young children. *Frontiers in Psychology*, *7*, Article 1181. <https://doi.org/10.3389/fpsyg.2016.01181>
- Gibson, E. J. (1969) *Principles of perceptual learning and development*. Prentice-Hall.
- Glidden, P. L. (2008). Prospective elementary teachers' understanding of order of operations. *School Science and Mathematics*, *108*(4), 130–136. <https://doi.org/10.1111/j.1949-8594.2008.tb17819.x>
- Goldstone, R. L., Marghetis, T., Weitnauer, E., Ottmar, E. R., & Landy, D. (2017). Adapting perception, action, and technology for mathematical reasoning. *Current Directions in Psychological Science*, *26*(5), 434–441. <https://doi.org/10.1177/0963721417704888>
- Gould, S. J. (1991). Exaptation: A crucial tool for an evolutionary psychology. *The Journal of Social Issues*, *47*(3), 43–65. <https://doi.org/10.1111/j.1540-4560.1991.tb01822.x>
- Gunnarsson, R., Sönnerhed, W. W., & Hernell, B. (2016). Does it help to use mathematically superfluous brackets when teaching the rules for the order of operations? *Educational Studies in Mathematics*, *92*(1), 91–105. <https://doi.org/10.1007/s10649-015-9667-2>
- Harrison, A., Smith, H., Hulse, T., & Ottmar, E. R. (2020). Spacing out! Manipulating spatial features in mathematical expressions affects performance. *Journal of Numerical Cognition*, *6*(2), 186–203. <https://doi.org/10.5964/jnc.v6i2.243>
- Hong, W., Star, J. R., Liu, R.-D., Jiang, R., & Fu, X. (2023). A systematic review of mathematical flexibility: Concepts, measurements, and related research. *Educational Psychology Review*, *35*(4), Article 104. <https://doi.org/10.1007/s10648-023-09825-2>
- Hornburg, C. B., Devlin, B. L., & McNeil, N. M. (2022). Earlier understanding of mathematical equivalence in elementary school predicts greater algebra readiness in middle school. *Journal of Educational Psychology*, *114*(3), 540–559. <https://doi.org/10.1037/edu0000683>
- Jiang, M. J., Cooper, J. L., & Alibali, M. W. (2014). Spatial factors influence arithmetic performance: The case of the minus sign. *Quarterly Journal of Experimental Psychology*, *67*(8), 1626–1642. <https://doi.org/10.1080/17470218.2014.898669>
- Kellman, P. J., & Massey, C. M. (2013). Perceptual learning, cognition, and expertise. In *Psychology of learning and motivation* (Vol. 58, pp. 117–165). Academic Press.

- Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science*, 2(2), 285–305. <https://doi.org/10.1111/j.1756-8765.2009.01053.x>
- Kieran, C. (1979). Children's operational thinking within the context of bracketing and the order of operations. In D. Tall (Ed.), *Proceedings of the 3rd Conference of the International Group for the Psychology of Mathematics Education* (pp. 128–133). Mathematics Education Research Centre, Warwick University, Coventry, United Kingdom.
- Landy, D., & Goldstone, R. L. (2007). Formal notations are diagrams: Evidence from a production task. *Memory & Cognition*, 35(8), 2033–2040. <https://doi.org/10.3758/BF03192935>
- Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *Quarterly Journal of Experimental Psychology*, 63(10), 1953–1968. <https://doi.org/10.1080/17470211003787619>
- Landy, D. H., Jones, M. N., & Goldstone, R. L. (2008). How the appearance of an operator affects its formal precedence. In *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society* (pp. 2109–2114). Cognitive Science Society.
- Lee, J.-E., Hornburg, C. B., Chan, J. Y.-C., & Ottmar, E. R. (2022). Perceptual and number effects on students' initial solution strategies in an interactive online mathematics game. *Journal of Numerical Cognition*, 8(1), 166–182. <https://doi.org/10.5964/jnc.8323>
- Lee, K., & Lee, H. W. (2019). Inhibition and mathematical performance: Poorly correlated, poorly measured, or poorly matched? *Child Development Perspectives*, 13(1), 28–33. <https://doi.org/10.1111/cdep.12304>
- Lincevski, L., & Livneh, D. (1999). Structure sense: The relationship between algebraic and numerical contexts. *Educational Studies in Mathematics*, 40(2), 173–196. <https://doi.org/10.1023/A:1003606308064>
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), Article 3139. <https://doi.org/10.21105/joss.03139>
- Maloney, E. A. (2016). Math anxiety: Causes, consequences, and remediation. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school* (2nd ed., pp. 408–423). Routledge.
- Marghetis, T., Landy, D., & Goldstone, R. L. (2016). Mastering algebra retrains the visual system to perceive hierarchical structure in equations. *Cognitive Research: Principles and Implications*, 1(1), Article 25. <https://doi.org/10.1186/s41235-016-0020-9>
- Mazzocco, M. M., Chan, J. Y.-C., & Bock, A. M. (2017). Early executive function and mathematics relations: Correlation does not ensure concordance. In C. Day-Hess, J. Sarama, D. Clements, & C. Germeroth (Eds.), *Advances in child development and behavior: The development of early childhood mathematics education* (pp. 290–307). Elsevier Academic Press. <https://doi.org/10.1016/bs.acdb.2017.05.001>
- McNeil, N. M. (2014). A change-resistance account of children's difficulties understanding mathematical equivalence. *Child Development Perspectives*, 8(1), 42–47. <https://doi.org/10.1111/cdep.12062>
- McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, 76(4), 883–899. <https://doi.org/10.1111/j.1467-8624.2005.00884.x>
- McNeil, N. M., Rittle-Johnson, B., Hattikudur, S., & Petersen, L. A. (2010). Continuity in representation between children and adults: Arithmetic knowledge hinders undergraduates' algebraic problem solving. *Journal of Cognition and Development*, 11(4), 437–457. <https://doi.org/10.1080/15248372.2010.516421>
- Medrano, J., & Prather, R. W. (2023). Rethinking executive functions in mathematical cognition. *Journal of Cognition and Development*, 24(2), 280–295. <https://doi.org/10.1080/15248372.2023.2172414>
- National Association of Education Progress, Question Tool. (2022). U.S. Department of Education. <https://www.nationsreportcard.gov/nqt/searchquestions>
- Ngo, V., Perez Lacera, L., Closser, A. H., & Ottmar, E. R. (2023). The effects of operator position and superfluous brackets on student performance in simple arithmetic. *Journal of Numerical Cognition*, 9(1), 107–128. <https://doi.org/10.5964/jnc.9535>
- Norton, S. J., & Cooper, T. J. (2001, August). Students' perceptions of the importance of closure in arithmetic: Implications for algebra. In *Proceedings of the International Conference of the Mathematics education into the 21st century project* (pp. 198–202). <https://sites.unipa.it/grim/ANortonCooper.PDF>
- Ottmar, E. R., Landy, D., Goldstone, R., & Weitnauer, E. (2015). Getting From Here to There! Testing the effectiveness of an interactive mathematics intervention embedding perceptual learning. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society* (pp. 1793–1798). Cognitive Science Society.

- Palmer, S. E. (1992). Modern theories of Gestalt perception. In G. W. Humphreys (Ed.), *Understanding vision: An interdisciplinary perspective* (pp. 39–70). Blackwell.
- Pillay, H., Wilss, L., & Boulton-Lewis, G. (1998). Sequential development of algebra knowledge: A cognitive analysis. *Mathematics Education Research Journal*, *10*(2), 87–102. <https://doi.org/10.1007/BF03217344>
- Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist*, *53*(3), 145–164. <https://doi.org/10.1080/00461520.2018.1447384>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rivera, J., & Garrigan, P. (2016). Persistent perceptual grouping effects in the evaluation of simple arithmetic expressions. *Memory & Cognition*, *44*(5), 750–761. <https://doi.org/10.3758/s13421-016-0593-z>
- Spielhagen, F. R. (2006). Closing the achievement gap in math: The long-term effects of eighth-grade algebra. *Journal of Advanced Academics*, *18*(1), 34–59. <https://doi.org/10.4219/jaa-2006-344>
- Spitzer, M. W. H., & Moeller, K. (2022). Predicting fraction and algebra achievements online: A large-scale longitudinal study using data from an online learning environment. *Journal of Computer Assisted Learning*, *38*(6), 1797–1806. <https://doi.org/10.1111/jcal.12721>
- Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology*, *40*, 41–54. <https://doi.org/10.1016/j.cedpsych.2014.05.005>
- Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, *18*(6), 565–579. <https://doi.org/10.1016/j.learninstruc.2007.09.018>
- Szokolszky, A., Read, C., Palatinus, Z., & Palatinus, K. (2019). Ecological approaches to perceptual learning: Learning to perceive and perceiving as learning. *Adaptive Behavior*, *27*(6), 363–388. <https://doi.org/10.1177/1059712319854687>
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & Von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. <https://doi.org/10.1037/a0029333>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, L. T., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. <https://doi.org/10.21105/joss.01686>
- Yu, J., Landy, D., & Goldstone, R. L. (2018). Visual flexibility in arithmetic expressions. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 2753–2758). Cognitive Science Society.
- Zazkis, R., & Rouleau, A. (2018). Order of operations: On convention and met-before acronyms. *Educational Studies in Mathematics*, *97*(2), 143–162. <https://doi.org/10.1007/s10649-017-9789-9>



*Journal of Numerical Cognition* (JNC) is an official journal of the Mathematical Cognition and Learning Society (MCLS).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.