The Effects of Perceptual and Embodied Features on Student Learning and Performance in Online Platforms

by

Avery Harrison Closser

A Dissertation

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in

Learning Sciences & Technologies

April 2022

APPROVED:

Erin Ottmar, Ph.D., Worcester Polytechnic Institute

Neil Heffernan, Ph.D., Worcester Polytechnic Institute

Jacob Whitehill, Ph.D., Worcester Polytechnic Institute

Mitchell J. Nathan, Ph.D., University of Wisconsin-Madison

Abstract

Thinking and learning are inherently tied to our perceptual processes and physical experiences in the world, yet this connection is typically underutilized in education and educational tools. As educational technologies are developed to support student learning, their design should be informed by theory and evidence to optimize the instructional support that students receive. The purpose of this work is to advance cognitive theories of learning and provide recommendations for researchers, teachers, and content developers to leverage students' perceptual processes and body-based resources in online instructional materials for math education. Specifically, this dissertation includes three studies that demonstrate how subtle perceptual and embodied features may be feasibly implemented in online instructional materials and how those features impact students' reasoning, performance, and learning in arithmetic and algebra. First, this dissertation describes the effects of spatial proximity between operands in order-of-operation problems on student performance. Second, this dissertation explores the relation between spatial proximity in notation, students' inhibitory control, and problem-solving performance. Finally, this dissertation describes how worked examples with different degrees of student interaction impact learning in online settings. Together, this body of work provides insights as to how cognitive theories may be leveraged in online learning environments by designing perceptual scaffolds and embodied features in instructional materials for math education.

Acknowledgments

This dissertation is a product of my five years at Worcester Polytechnic Institute (WPI), including the skills acquired and insights learned as I earned my Master's degree, pursued my Ph.D., and gained experiences to support my career goals. I am so grateful for the individuals listed below as well as everyone I have been fortunate enough to learn from, work with, and laugh with over the past five years in the WPI community and through external collaborations.

Thank you to my committee members: Erin Ottmar, Neil Heffernan, Jacob Whitehill, and Mitchell Nathan. Your mentorship, support, and advocacy have been instrumental in my training.

Thank you to my mentor and collaborator, Jenny Chan. Thank you for your mentorship and kindness and for sharing your project management and writing practices with me.

Thank you to my lab members, collaborators, and mentors across the Learning Sciences & Technologies community. In particular, thank you to Anthony Botelho, Hannah Smith, Taylyn Hulse, Dan Manzo, John Erickson, Vy Ngo, and Luisa Perez Lacera.

Thank you to the teachers and students that made this research possible. Although the past couple of years have been particularly difficult, you found a way to participate in these studies. Thank you for your commitment to advancing educational practices.

Thank you to my family members and friends for your love, support, and interest in my work. The support system you provide is the key to any success I have and lots of my happiness.

Most importantly, thank you to my wonderful husband, Cole, for your love, support, and undying patience. As proud as I am of the research presented here, I am overwhelmed with joy by the milestones that we have reached together during the past five years outside of work. Thank you for marrying me, moving to new states, raising a puppy, and making life as sweet as possible. To our next adventure!

Funding

The research presented in this dissertation has been financially supported by different funding avenues. I was awarded a Marietta Anderson Graduate Fellowship by Worcester Polytechnic Institute (WPI) to cover my tuition and research stipend for the 2017-2018 school year. I was supported by the Graduate Assistance in Areas of National Need in 2018-2019 under fellowships #P200A150306 and #PA200A180088 awarded to WPI. Following that, I was awarded a National Science Foundation Graduate Research Fellowship (#1645629) to cover my tuition and research stipend from 2019-2022. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. Additionally, I was awarded a Psi Chi (The International Honor Society in Psychology) Graduate Research Grant for 2021-2022 to support the project presented in Chapter 4. Any funding that has supported the educational technology platforms used for this research is acknowledged within the relevant chapter.

Table of Contents

Abstract 1

Acknowledgments 2

Funding 3

Table of Contents 4

Chapter 1. Introduction 8

Study 1: The Effect of Spacing in Math Expressions on Student Performance 10

Study 2: The Relation Between Spacing, Inhibition Skills, and Math Performance 11

Study 3: The Impact of Action and Self-Explanation in Worked Examples on Learning 12

References 14

Chapter 2. Spacing Out! Manipulating Spatial Features in Mathematical Expressions Affects Performance 16

Abstract 17

Spacing Out! Manipulating Spatial Features in Mathematical Expressions Affects Performance 18

The Present Study 22

Methods 23

Study Context 23

Participants 24

Experimental Conditions and Procedure 25

Measures 28

Prior Mathematics Performance 28

Grade Level 29

Approach to Analysis 29

Results 30

Descriptive Statistics 30

Hierarchical Linear Models Examining the Impact of Spacing Condition on Mastery Speed 32

Model 3: Research Question 1: Does Spacing Influence Assignment Mastery Speed? 33

Model 4: Research Question 2: Does Prior Performance Moderate the Effect of Condition on Assignment Mastery Speed? 34

Model 5: Research Question 3: Does Grade Moderate the Effect of Condition on Assignment Mastery Speed? 35

Discussion 35

Physical Spacing in Math Expressions Affects Student Performance 36
Prior Math Performance, Grade Level, and Spacing Conditions 38
Implications for Teaching and Learning 39
Limitations 40
Future Directions 40
Conclusion 42
Funding 43
Competing Interests 43
Acknowledgements 43
References 44

Chapter 3. Resisting the Urge to Calculate: The Relation Between Inhibition Skills and Perceptual Cues in Arithmetic Performance 48
Abstract 49
Resisting the Urge to Calculate: The Relation Between Inhibitory Control and Perceptual Cues in Arithmetic Performance 50
Perceptual Learning: Theory and Mechanisms 51
Spacing as a Perceptual Grouping Mechanism 53
Inhibition and Mathematics 54
Current Study 56
Methods 57
Participants 57
Study Design and Procedure 58
Materials 59
Order-of-Operations Problems 59
Stroop Task 60
Measures 61
Approach to Analysis 63
Results 65
Preliminary Analysis 65
Effects of Physical Spacing on Congruent and Incongruent Problem Solving 68
Pre-registered Analysis: Effect of Perceptual Cues Controlling for Baseline Accuracy 68
Exploratory Analysis: Effect of Perceptual Cues by Baseline Performance 69
Effects of Inhibitory Control on Order of Operations Problems 71
Pre-registered Analysis: Effects of Inhibitory Control with the Baseline Covariate 71
Exploratory Analysis: Effects of Perceptual Cues by Inhibitory Control 72
Discussion 74
Effect of Physical Spacing on Performance 75

Limitations and Future Directions 78
Conclusions 79
References 80

Chapter 4. Viewing vs. Mirroring: The Effects of Action and Self-Explanation in Worked Examples on Algebra Learning 86
Abstract 87
Viewing vs. Mirroring: The Effects of Action and Self-Explanation in Worked Examples on Algebra Learning 88
Theoretical Background 90
Cognitive Load Theory 90
Embodied Cognition and Design 91
Testing and Integrating Cognitive Theories to Design Worked Examples 98
The Current Study 99
Methods 101
Participants 101
Study Procedure 102
Day One: Pretest 103
Algebra Knowledge Assessments 103
Day Two: Intervention 104
Tutorials 104
Worked Examples and Paired Practice Problems 105
Conditions 107
Cognitive Load Measure 111
Day Three: Posttest 111
Approach to Analysis 112
Results 113
Preliminary Analyses 113
Primary Analysis 115
Effects of Worked Example Condition on Cognitive Load 116
Discussion 118
The Worked Example Effect 118
Viewing vs. Mirroring Worked Examples 120
The Impact of Self-Explanation Prompts 121
The Impact of Worked Examples on Cognitive Load 122
Limitations and Future Directions 123
Implications and Conclusions 126
References 128

Chapter 5. Discussion 142

Summary and Interpretations of Findings 142
Current Limitations Invite Future Directions 144
    Embracing Complexity with Worked Examples and Instructional Materials 145
    Implementing Long-Term Instructional Support in Online Platforms 146
    Advancing Methodological Practices in the Learning Sciences 147
Conclusion 149
References 151

Chapter 1. Introduction

Students' mathematics skills, education, and achievement are related to multiple outcomes throughout their lifespan. For instance, childhood performance in mathematics as early as kindergarten is a predictor of later achievement (Claessens & Engel, 2013). Further, mathematics education and achievement are related to neuroplasticity and brain development during adolescence (Zacharopoulos et al., 2021), and employment during adulthood (Parsons & Bynner, 2005). Importantly, algebra is commonly considered a gateway to higher-level mathematics for students in high school and beyond (Matthews & Farmer, 2008). In particular, students who progress beyond Algebra 1 are more likely to outperform their peers on national assessments (Kena et al., 2015). However, many students struggle to grasp basic algebraic concepts, barring most students from pursuing further mathematics (Kena et al., 2015). This struggle may be partially attributed to the challenge that students face in learning procedural rules as well as the conceptual knowledge to appropriately apply those rules in practice. Therefore, it is critical to provide effective instructional support for students to acquire procedural skills and conceptual knowledge necessary for success in algebra.

I posit that by harnessing the knowledge that cognition is shaped by our physical experiences, learning technologies and tools can design more effective instructional support (e.g., through perceptual scaffolding and worked examples) to help students progress beyond early algebra. My theoretical framework is largely informed by theories of perceptual learning (e.g., Closser et al., under review; Goldstone 1998; Goldstone et al., 2017) and embodied cognition and learning (e.g., Nathan, 2014, 2021). Together, these theoretical perspectives argue that thinking and learning are reflected in, and at least partially impacted by, a cyclical relationship between: a) our perceptual processes which interpret *incoming* information from the environment, and b) our physical actions which provide *outgoing* displays of cognition and garner feedback from our environments. In terms of math education, understanding how students' perceptual processes and body-based resources (e.g., actions, movement, speech, gestures) influence reasoning and learning may present new opportunities in instructional practice and content design in educational technologies to support student thinking and learning in math.

This dissertation aims to demonstrate: a) how perceptual scaffolds and embodied features may be feasibly implemented in online activities across two learning environments, and b) how these subtle features impact students' performance and learning in algebra. Importantly, this research is not testing the efficacy of any learning technologies. Instead, multiple learning technologies have been used as research platforms to tease apart how multiple cognitive mechanisms may work together to influence learning. Through this line of research, I aim to inform cognitive theories of learning as well as improve online and technology-augmented learning environments by providing recommendations for researchers, teachers, and content developers that leverage our body-based resources for learning. To reflect this work, I present three projects that demonstrate how perceptual and embodied scaffolds may be implemented in online learning systems and investigate how these subtle features in online instructional materials may shape students' reasoning, learning, and performance in math. This research was conducted across two online learning platforms (i.e., ASSISTments and Graspable Math) as well as through two additional research platforms (i.e., Qualtrics and Psychopy). First, I include a published manuscript detailing the effects of spatial proximity between operands in order-of-operation problems on student performance (Harrison et al., 2020). Second, I include an in-preparation manuscript that extends this work by exploring the relation between spatial proximity in notation, students' inhibitory control, and problem-solving performance. Finally, I present preliminary results from my proposed study: comparing how worked examples impact student learning in online settings when they are designed from principles of cognitive load theory and embodied cognition. Specifically, these studies address:

1) *How do perceptual cues (i.e., spacing) in arithmetic problems impact students' performance?*
2) *How do perceptual cues (i.e., spacing) in arithmetic problems interact with students' inhibition skills to impact performance?*
3) *How do features of worked examples that leverage student interaction impact learning in algebra?*

I close the dissertation with a "Discussion" in which I describe key takeaways and insights, address limitations, and share future avenues for advancing cognitive theories of learning, informing instructional practice, and providing recommendations for educational technology design. A synopsis of each project is presented below.

Study 1: The Effect of Spacing in Math Expressions on Student Performance

In Chapter 2, I present the manuscript, "*Spacing Out: Manipulating Spatial Features in Mathematical Expressions Affects Performance*" (Harrison et al., 2020), which was published in the *Journal of Numerical Cognition.* This experiment provides confirmatory evidence of the effect of spacing between symbols within mathematical expressions on student performance. Specifically, we investigated how the presentation of expressions impacted students' adherence to the order of operations (e.g., performing calculations within parentheses first; completing multiplication and division calculations before addition and subtraction). A total of 2,152 students in fifth through twelfth grade were randomly assigned to one of four conditions within an online problem set, with symbols in the algebraic expressions spaced 1) *neutrally*, with no spaces in the expression (e.g., 4*5+3), 2) *congruent* with the order of operations through grouping terms (e.g., 4*5 + 3), 3) *incongruent* with the order of operations (e.g., 4 * 5+3), or 4) *mixed*, a combination of the previous conditions. We found that students who viewed problems with incongruent spacing made more errors and had to solve more problems to complete the assignment than those who viewed congruent or neutrally spaced problems. Additionally, students who viewed problems with mixed spacing had to solve more problems to complete the assignment than students who viewed congruent problems. We concluded that viewing expressions with spacing that is incongruent with the order of operations presents challenges for students. Overall, these results replicated prior research on perceptual learning (e.g., Landy & Goldstone, 2010) in an online homework environment and support the claim that spacing between symbols influences student performance on order-of-operations problems. From these findings, we contend that online platforms could leverage spacing to help students learn and attend to the order of operations then fade perceptual support over time.

Study 2: The Relation Between Spacing, Inhibition Skills, and Math Performance

The previous study (Harrison et al., 2020) demonstrates how robust the impact of perceptual cues on students' problem-solving performance can be, extending across grade levels and prior knowledge. However, other cognitive skills may interact with perceptual cues to impact students' problem-solving performance. Prior work has shown the relation between executive function and mathematics performance and specifically, between inhibition skills (i.e., the ability to suppress prepotent responses) and problem solving (Cassotti et al., 2016; Cragg & Gilmore, 2014). In Chapter 3, I present an online experiment in which college students completed a modified version of the Stroop task followed by order-of-operations problems presented with neutral, congruent, and incongruent spacing. I predicted that students with stronger inhibition skills may be better able to suppress the urge to calculate invalid solutions primed by incongruent spacing. Results showed that, controlling for students' baseline accuracy on neutral problems, there were no significant differences in students' accuracy when solving congruent vs. incongruent order-of-operations problems. However, students had significantly longer response times on congruent as opposed to incongruent problems. There were no main effects of inhibitory control on problem-solving performance. These results advance perceptual learning theory by exploring the impact of spatial proximity on students' performance when accounting for baseline performance and add to the growing debate of whether and how inhibitory control may be associated with mathematics skills and performance.

Study 3: The Impact of Action and Self-Explanation in Worked Examples on Learning

While the previous studies investigated the influence of perceptual cues on students' performance *during* an assignment, this project examines how theories of embodied cognition may inform the design of worked examples for online platforms to support students as they are learning new concepts and procedural rules *prior* to problem-solving practice. In Chapter 4, I present my research investigating how effectively students learn from worked examples that vary in degree of interaction (i.e., with vs. without self-explanations) and embodiment (i.e., viewing or mirroring worked examples on-screen). While worked examples have largely been designed from cognitive load theory to offload strains on students' cognitive capacities and free up working memory to support learning (Chandler & Sweller, 1991), limited work has explored how to leverage other theories in the design of online worked examples. Drawing from theories of embodied learning and design (e.g., Abrahamson et al., 2020; Nathan, 2014), I leveraged the affordances of a dynamic algebra notation tool (Graspable Math) to test how student actions and self-explanations impact learning from different worked example formats.

I predicted that algebra students may learn more when they dynamically *mirror* worked examples on-screen, rather than simply *view* worked examples. To test this hypothesis, I designed and implemented an online RCT to compare the effects of viewing versus mirroring worked examples on learning. A total of 64 ninth-grade Algebra I students completed an online, three-day study in which they were randomly assigned to: a) *view,* b) *view-and-explain*, c) *mirror*, or d) *mirror-and-explain* worked examples and complete paired practice problems. Chapter 4 presents the preliminary results from this study. Namely, all students improved from pretest to posttest with no significant differences in students' learning gains between those who viewed vs. mirrored worked examples. However, students who received self-explanation prompts with their worked examples did learn more. These preliminary results support prior research on the worked example effect as well as the value of self-explanation prompts for student learning. With a full sample, the findings from this study should advance cognitive theories and provide recommendations for how worked examples can and should be effectively designed for online learning environments.

References

Abrahamson, D., Nathan, M. J., Williams-Pierce, C., Walkington, C., Ottmar, E. R., Soto, H., & Alibali, M. W. (2020, August). The future of embodied design for mathematics teaching and learning. In *Frontiers in Education* (Vol. 5, p. 147). Frontiers.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24-34.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293-332.

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, *115*(6), 1-29.

Closser, A. H., Chan, J. Y. C., Smith, H., & Ottmar, E. (under review). Perceptual learning in math: Implications for educational research, practice, and technology. Rapid Community Report Series. Digital Promise and the International Society of the Learning Sciences.

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.

Harrison, A., Smith, H., Hulse, T., & Ottmar, E. (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

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., ... & Velez, E. D. V. (2015). The Condition of Education 2015. NCES 2015-144. *National Center for Education Statistics*.

Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *The Quarterly Journal of Experimental Psychology*, *63*(10), 1953-1968.

Matthews, M. S., & Farmer, J. L. (2008). Factors affecting the algebra I achievement of academically talented learners. *Journal of Advanced Academics*, *19*(3), 472-501.

Nathan, M. J. (2014). Grounded mathematical reasoning. In L. Shapiro (Ed.). *The Routledge handbook of embodied cognition* (pp. 171-183). Albingdon, UK: Routledge.

Nathan, M. J. (2021). *Foundations of embodied learning: A paradigm for education*. Routledge.

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* London: National Research and Development Centre for Adult Literacy and Numeracy.

Zacharopoulos, G., Sella, F., & Kadosh, R. C. (2021). The impact of a lack of mathematical education on brain development and future attainment. *Proceedings of the National Academy of Sciences*, *118*(24).

Chapter 2. Spacing Out! Manipulating Spatial Features in Mathematical Expressions Affects Performance

This chapter presents the pre-print version of the following manuscript:

Harrison, A., Smith, H., Hulse, T., & Ottmar, E. (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

Abstract

The current study explores the effects of physical spacing within mathematical expressions on student performance. A total of 2,152 students in 5th-12th grade were randomly assigned to one of four conditions within an online problem set, with terms in algebraic expressions spaced 1) *neutrally,* with no spaces in the expression, 2) *congruent* with the order of precedence through grouping terms, 3) *incongruent* with the order of precedence, or 4) *mixed*, a combination of the previous conditions. Results show that students who viewed incongruent problems made more errors and had to solve more problems to complete the assignment than those who viewed congruent or neutrally spaced problems. Additionally, students who viewed problems with mixed spacing had to solve more problems to complete the assignment than students who viewed congruent problems. These findings suggest that viewing expressions with spacing that is incongruent with the order of precedence presents challenges for students. Overall, these results replicate prior research in perceptual learning in a natural homework environment and support the claim that physical spacing between terms does influence student performance on order of precedence problems.

Keywords: perceptual learning, spatial proximity, mathematical cognition, mathematical operations

Spacing Out! Manipulating Spatial Features in Mathematical Expressions Affects Performance

Formal mathematics is a commonly used example of how humans make sense of abstract symbolic reasoning (Anderson, 2007; Goldstone et al., 2017). However, learning mathematics is difficult for many students, in part because of the requirement to learn and execute abstract rules as they apply to mathematical notation. Being able to interpret symbolic notation and compute simple calculations efficiently and accurately is critical for solving more complex mathematics problems, notably algebra. For example, the order of precedence stipulates how to simplify an expression or equation, including the order in which computations can be carried out. Such abstract rules require students to learn seemingly arbitrary conceptual processes and appropriately apply those rules when reasoning about mathematics.

Beyond being abstract and requiring conceptual knowledge, reasoning about mathematics is also inherently perceptual (Marghetis et al., 2016) with ample evidence suggesting that mathematical processing and understanding is influenced by the visual presentation of mathematical notation (McNeil & Alibali, 2004, 2005). Perceptual learning has been suggested to be a mechanism that adapts perception and directs attention to relevant information in the environment (Gibson, 1969), supporting high level cognition. For instance, spatial proximity, a Gestalt law that posits that individuals perceive objects in close proximity to be a group, has been shown to bias mathematical reasoning. This phenomena in mathematics supports the notion that people rely on perceptual cues to process symbolic notations and are heavily influenced by spatial properties of notation (Goldstone et al., 2017; Wagemans et al., 2012). Regardless of conceptual knowledge, the tendency to use perceptual cues and groupings in mathematics notation is somewhat automatic and has implications for the ways in which individuals interpret, compute, and produce mathematics notation.

Although subtle visual manipulations are irrelevant to the mathematical meaning of notation, visual manipulations of notation can lead to attentional biases and create perceptual groupings among terms and operands. For instance, terms and operands spaced in close proximity within a mathematical notation tend to be seen as a group, such as viewing "4 x 6+3" and wrongly grouping "6+3" together based on the spatial proximity of those terms (Jiang, Cooper, & Alibali, 2014; Kirshner, 1989; Landy & Goldstone, 2007b, 2010). Additionally, while novice learners often solve order of precedence problems based on memorized rules, experts have been shown to rely on perceptual cues when solving complex equations (Braithwaite et al., 2016; Rumelhart et al., 1986), providing evidence that there may be a shift at some point in experience or procedural fluency from attending to abstract rules of formal mathematics to attending to perceptual cues in formal mathematical notation.

A large body of research has demonstrated that the physical spacing between terms and operands within equations and expressions contributes to students' perceptions of *how* they are able, within the rules of mathematics, to interpret meaning and perform computations (e.g., Jiang et al., 2014; Landy & Goldstone, 2007a, 2007b, 2010; Rivera & Garrigan, 2016). Consequently, spacing in mathematical expressions has been found to impact performance on equation-solving. For instance, Landy and Goldstone (2007b, 2010) manipulated whether the spacing of terms in expressions was congruent (multiplications spaced closer than additions) or incongruent (additions spaced closer than multiplications) with the order of precedence. They found that participants made more errors and were more likely to perform addition before multiplication in the incongruent spacing condition. For example, in the case of 7+1 * 4, people often first combine the 7+1 to make 8 and then multiply by 4 to get 32, instead of properly multiplying 1 by 4 and then adding 7. These results suggest that mathematical reasoning is at least somewhat perceptually driven through low-level visual and attentional factors. Landy and Goldstone (2007b, 2010) posit that this effect occurs because spacing cues bias individuals to perform specific operations, even if those cues are mathematically invalid.

Similarly, Jiang and colleagues (2013) found that when participants viewed operand spacing in expressions which created perceptual groupings incongruent with the order of precedence, participants tended to make target errors reflective of incorrectly grouping a set of terms. Rivera & Garrigan (2016) extended this work by replicating the effect of incongruent spacing on order of precedence errors found by Landy and Goldstone (2010), providing further support for the effects of perceptual grouping on mental arithmetic, even in the case of evaluating simple expressions. This work provides evidence that when perceptual grouping is incongruent with operator precedence, the likelihood of order of precedence errors in mental arithmetic increases. More broadly, this research shows that individuals use perceptual spacing to interpret and reason about mathematics and may have a difficult time ignoring perceptual cues even if they are incongruent with mathematical rules.

Gómez, Benavides-Varela, Picciano, Semenza, and Dartnell (2014) extended this work with a sample of 5^{th}-8^{th} grade Chilean and Italian students and found that the spacing effects seem to emerge in younger students as well. Braithwaite et al. (2016) also explored the effect of physical spacing outside of a laboratory setting among even younger primary-school children (equivalent to U.S. grade levels 2-6) in the Netherlands and found higher error rates for individuals who viewed problems which had spacing incongruent with the order of precedence. They also found that this effect of spacing increased with grade level, further suggesting that there is an increased reliance on perceptual grouping with age and experience with arithmetic. However, aside from the work of Gómez et al. (2014) and Braithwaite et al. (2016), less is known about whether or how perceptual grouping, influenced by physical spacing within mathematical expressions, impacts student behavior in typical school settings and varies across grade levels.

Overall, this body of literature shows that the perceptual grouping of mathematical terms in an expression or equation influences both novices and experts during problem solving. Specifically, when terms are spatially organized in groups that mirror the order of precedence, students are more likely to have higher performance (Landy & Goldstone, 2007a) and more accurate interpretations (Jiang et al., 2014; Landy & Goldstone, 2010). Conversely, when terms are grouped in ways that are incongruent with the order of precedence, students are more likely to take more time to solve (Gómez, Bossi, & Dartnell, 2014) and make more errors (Jiang et al., 2013; Landy & Goldstone, 2007, 2010; Rivera & Garrigan, 2016). Such research provides evidence of the influence of perceptual learning on mathematical problem solving, which could play a key role in student learning. To further this area of research, the current study aims to replicate and extend prior research by exploring the effects of spacing on student performance on order of operations problems with upper elementary through high school algebra learners in a natural homework setting.

The present study asked 5-12th grade students to simplify order of precedence expressions in ASSISTments, an online tutoring system (Heffernan & Heffernan, 2014). Students were randomized into one of four experimental conditions, which manipulated the physical spacing between numbers and terms within mathematical expressions to be either *neutral*, *congruent* or *incongruent* with the order of precedence, or a *mixed* combination of spacing across problems. We then examined whether there were differences in content mastery speeds (the total number of problems that students had attempted by the time that they correctly answered three problems in a row) based on spacing conditions.

This study extends prior research on perceptual learning in four key ways. First, the majority of studies examining the effects of physical spacing between mathematical terms have been conducted with undergraduate students in controlled laboratory settings rather than with school-aged children in authentic classroom and learning contexts. Second, the study is conducted through an online homework assignment assigned to students by their teachers using the ASSISTments platform (Heffernan & Heffernan, 2014), rather than administered by a researcher in a laboratory setting. Third, this study examines the effect of a mixed condition, where students are exposed to each of the experimental spacing conditions. Lastly, while many studies on perceptual learning have used error rates as the learning outcome, the current study uses mastery speeds, a measure of the number of problems attempted to master the material presented in the assignment as the dependent measures. From this extension of related research, this project aims to contribute a richer understanding of how perceptual grouping, from physical spacing in mathematical expressions, affects students' behavior in authentic learning contexts.

The Present Study

To extend prior research on perceptual learning as it pertains to mathematics performance, we present a randomized controlled trial with upper elementary, middle, and high school students in ASSISTments, an online tutoring system. This study is designed to explore the impact of physical spacing between terms on students' mastery speeds when solving a series of order of operations problems. Specifically, We posed the following questions:

1. *Does spacing impact assignment mastery speed?* We hypothesize that students who view congruent or neutrally spaced problems will have quicker mastery speeds (attempting fewer problems before correctly answering three problems in a row) compared to students who view incongruent spacing or mixed spacing problems.
2. If there are differences in assignment mastery speed based on condition, *does student prior performance moderate the relationship between condition and mastery speed?* We explore possible interactions between condition and prior performance to see if different levels of prior performance heighten or mitigate the effect of any spacing condition(s) on mastery speeds.
3. If there are differences in assignment mastery speed based on condition, *does grade level moderate the relationship between condition and mastery speed?* We explore possible interactions between condition and grade to see if the effect of any spacing condition(s) on mastery speed varies by grade level.

Methods

Study Context

Data for this study was collected from 2015-2019 in ASSISTments, an online tutoring system that features free content for K-12 students with a primary focus on mathematics (Heffernan & Heffernan, 2014). In addition to providing a technology tool for teachers to assign content and homework to students, ASSISTments also provides researchers with an experimental platform where independent researchers can create their own randomized controlled trials to be used by teachers and students. The de-identified data used in this analysis is available on Open

Science Framework (10.17605/OSF.IO/BAEUJ). Additionally, the original data report from ASSISTments is available at tiny.cc/spacingdata for further reference.

This randomized controlled trial was created by the authors and deployed as an available Skill Builder problem set covering order of operations content (targeting 7th grade) within ASSISTments. "Skill Builders" are optional problem sets that teachers can assign to provide students with fluency practice on topics commonly featured on standardized mathematics tests. Skill Builders map onto content areas from the Common Core State Standards and present problems from a given content area in a randomized order. These problem sets are designed to challenge a student in a mathematics topic until that student achieves content mastery.

Under default settings, students must consecutively answer three problems in a row correctly to achieve mastery status for the Skill Builder assignment. If a student answers a problem incorrectly, the problem count restarts and they continue to receive problems until they correctly answer three problems in a row. Therefore, in this context, a slower mastery speed (solving more problems in order to get three problems correct in a row) is an indicator of higher error and lower mathematics performance on a Skill Builder assignment. Mastery speed has been used as an outcome measure of student performance in previous ASSISTments studies (e.g., Botelho et al., 2015).

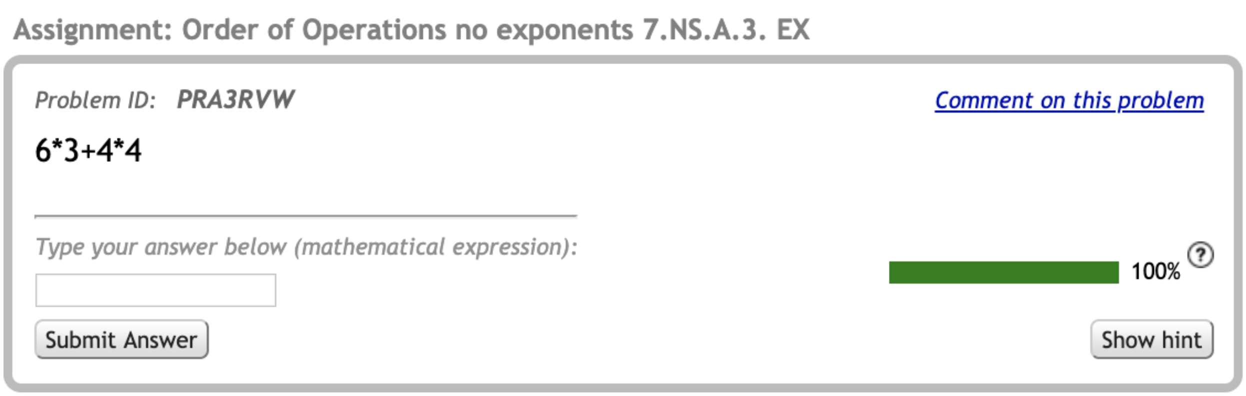
Participants

The final sample included in the analyses were 2,152 students (48.0% male, 35.2% female, 16.9% unknown) who completed more than three problems in the Skill Builder problem set and completed the assignment by achieving mastery. Participants were 5th-12th grade students assigned to complete the given problem set by their classroom teacher. The 2,152 students included in the final sample from this study came from 199 classes taught by 115 teachers from 83 schools in 64 districts from 16 states. The students were distributed across several grade levels, with a majority of students in middle school classrooms (.6% fifth, 11.7% sixth, 30.4% seventh, 9.8% eighth, 18.5% ninth, 1% tenth, .1% eleventh, and 1.1% of reported cases in twelfth grade; with the remaining 26.9% of cases missing grade level information).

Many more students initially opened the problem set but were dropped from this study for the following reasons. A total of 6,238 students opened the problem set, however, 4,053 students were excluded due to assignment completion within three problems or stopping the assignment within the first three problems, thus never seeing an experimental condition. Additionally, a small subset of participants was also excluded due to having an unknown mastery status for the problem set (*n* = 33).

Figure 1

Problem with Neutral Spacing as Shown in First Three Assignment Problems



Experimental Conditions and Procedure

When students opened the problem set, they were first exposed to three neutrally spaced expressions to solve (Figure 1). After completing the three neutrally-spaced problems, students were randomly assigned to one of four spacing conditions: 1) *neutral* (*n* = 574), with no spaces in the expression, 2) *congruent* (*n* = 555), with spacing which follows the order of precedence, 3) *incongruent* (n = 493; see Figure 2), with spacing which does not follow the order of precedence, or 4) *mixed* (n = 530), a combination of the previous conditions. Once assigned to a condition, students were presented with additional problems to solve. The problems in each condition were identical in structure but varied in the physical spacing of terms within each expression. The first several problems for each condition are shown in Table 1.

Table 1

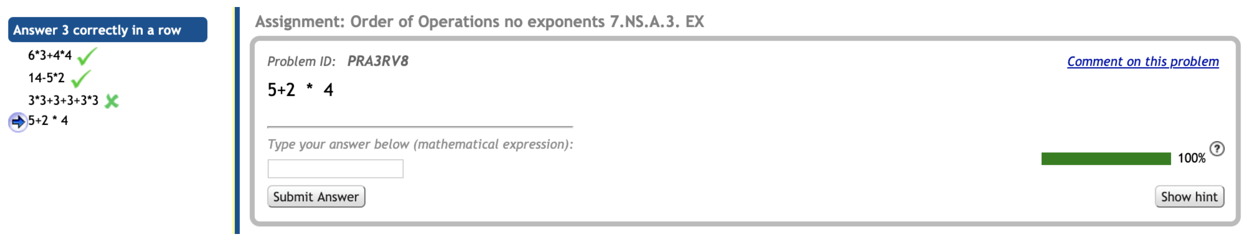
The First Eight Problems Assigned by Condition

	Tutorial: All Participants			
1.	6*3+4*4			
2.	14-5*2			
3.	3*3+3+3*3			
	Neutral	Congruent	Incongruent	Mixed
4.	5+2*4	5 + 2*4	5+2 * 4	5 + 2*4
5.	7*2+8*5	7*2 + 8*5	7 * 2+8 * 5	7 * 2+8 * 5
6.	4*3+2	4*3 + 2	4 * 3+2	4*3+2
7.	4*(2+5)+12-2*3	4*(2 + 5) + 12 - 2*3	4 * (2+5)+12-2 * 3	4*(2 + 5) + 12 - 2*3
8.	5+3*2	5 + 3*2	5+3 * 2	5+3 * 2

Most students continued to solve problems until they achieved mastery (answering three consecutive problems correctly on the first try). However, if a student answered a problem incorrectly, they could not move on to the next problem until typing in the correct answer. To support students as they moved through the assignment, one hint restating the order of operations was available to click on at the beginning of each problem (Figure 3). If students elected to see the hint, they were then immediately able to click “Show Answer” which would display the correct answer to type as the solution. Importantly, if students opted to view the hint, the problem was marked as incorrect and did not count towards the three mastery problems required to complete the assignment. Additionally, students could opt to stop the assignment at any time without completion. However, for students that did achieve mastery status, ASSISTments automatically closed the assignment and marked the status as completed.

Figure 2

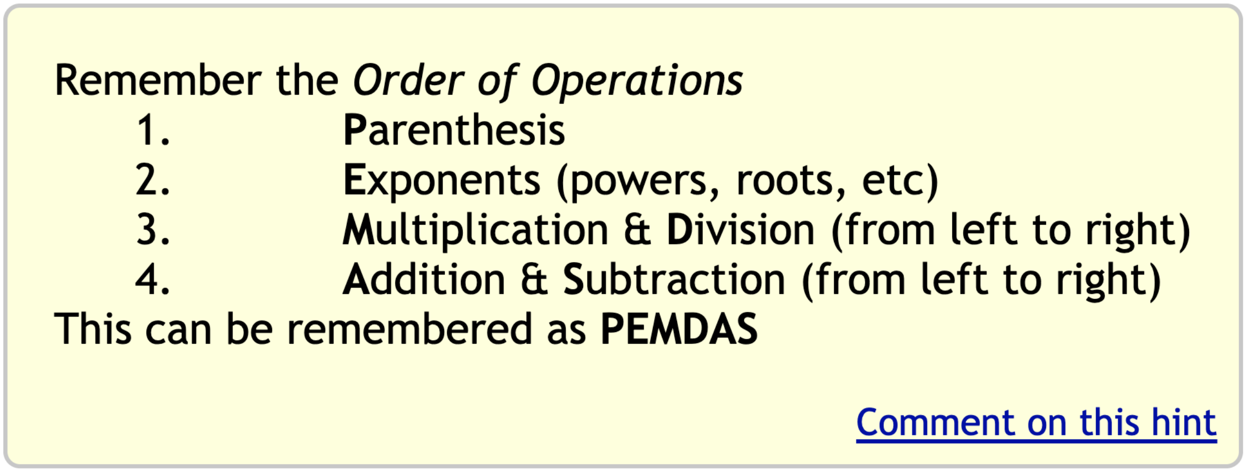
Example Assignment Screen for a Participant in the Incongruent Spacing Condition



The study remained open as an active Skill Builder for the order of operations standard without exponents (Common Core Standard 7.NS.A.3. EX) that teachers could easily assign to their students at any time for three years. At the end of the three years, the data from the study was aggregated using the ASSISTments Assessment of Learning Infrastructure (ALI) report that was automatically generated by the ASSISTments team for external researchers and provides aggregated data files at various levels of granularity such as student-level and problem-level (Ostrow et al., 2016). All variables of interest in this study were extracted from this report and are described in more detail below.

Figure 3

Hint Available to Participants on Each Problem in the Assignment



Measures

Prior to data analysis, the following measures for analyses were defined and extracted from the ASSISTments report as necessary for analysis.

Prior Mathematics Performance

As an estimated measure of prior mathematics performance, ASSISTments calculates a prior proportion correct value (from 0-1). This value represents the proportion of all previous ASSISTments problems completed from other assignments that each student answered correctly prior to the current experiment. However, the type of content may have varied and some participants may have had extensive experience with ASSISTments over years whereas others might have been first- or second-time users. Although participants varied in previous exposure and practice with ASSISTments, this value serves as a proxy for prior mathematics performance and has been used in studies that were deployed using the ASSISTments platform (e.g., Walkington et al., 2019). The distribution of prior performance scores was bimodal only due to a small subset (3.5%) of students who had demonstrated perfect prior performance in ASSISTments. This value was used as a continuous covariate for prior performance in analyses.

Grade Level

The ASSISTments ALI report also provided an ordinal value representing the reported grade level of each participant by the classroom teacher. With values ranging from 5-12, grade level was treated as a continuous variable in all analyses.

Approach to Analysis

After preprocessing the data, descriptive statistics were calculated in SPSS to determine means and variability for each variable and relations between each construct. Next, we conducted a one-way analysis of variance (ANOVA) with condition (neutral, congruent, incongruent, and mixed) as the independent variable and mastery speed as the dependent measure. We also conducted post hoc tests with Bonferroni correction to examine where there were significant differences in average mastery speed between conditions.

In addition to the ANOVA, we examined the impact of condition, above and beyond prior performance and grade level. To determine whether or not multilevel analysis would be appropriate, we calculated the intraclass correlation coefficient (ICC) from an unconditional 2-level hierarchical linear model (HLM; Model 1). An unconditional HLM model predicting mastery speed suggested that approximately 10% of the variance in mastery speed was attributable to differences at the class level. As this value exceeds the 7% variance threshold to suggest that using HLM would be appropriate (Lee, 2010; Niehaus et al., 2014), we chose to use HLM for all analyses to account for the nesting of students in classes.

Next, four two-level HLMs were conducted to explore our research questions. Model 2 estimates how the covariates, grade level and prior performance, impact participants' assignment mastery speed while accounting for any nested effects between the student and class levels. Model 3 includes the three condition variables (neutral, congruent, and mixed, with incongruent as the reference group) and estimates how the physical spacing between terms impacts participants' assignment mastery speed (compared to the incongruent condition) while accounting for any nested effects between the student and class levels.

Model 3 in HLM has the following form:

$$MASTERYSPEED_{ij} = \gamma_{00} + \gamma_{10}(CLASSGRADE_{ij}) + \gamma_{20}(NEUTRAL_{ij}) + \gamma_{30}(CONGRUENT_{ij}) + \gamma_{40}(MIXED_{ij}) + \gamma_{50}(PRIORPERC_{ij}) + \mu 0_j + r_{ij}$$

where *i* is students 1 through n, and *j* is class 1 through n.

Interaction terms were created and added to the hierarchical linear model to examine interactions between *prior performance and condition* as well as *grade level and condition* as predictors of mastery speed, controlling for prior performance and grade level as covariates. Model 4 presents results for the second research question, exploring whether an interaction between prior performance and spacing condition predicted mastery speed. Lastly, Model 5 presents results for the third research question, which explores whether an interaction between grade level and spacing condition predicted mastery speed.

Results

Descriptive Statistics

Overall, all students completed the assignment by eventually achieving mastery status (answering three problems correctly in a row) at some point in the assignment (*M*=6.38 problems, *SD*=3.24 problems). While working on the problem set, 33.6% of participants used the available hint at least once. See Table 2, below, for details on students' prior performance and average mastery speed by grade level. Prior performance scores indicated that, on average, students had correctly answered 70% of previously attempted problems in ASSISTments prior to the beginning of this study (*M* = .70, *SD* = .14).

Table 2

Descriptive Statistics on Student Performance by Grade Level

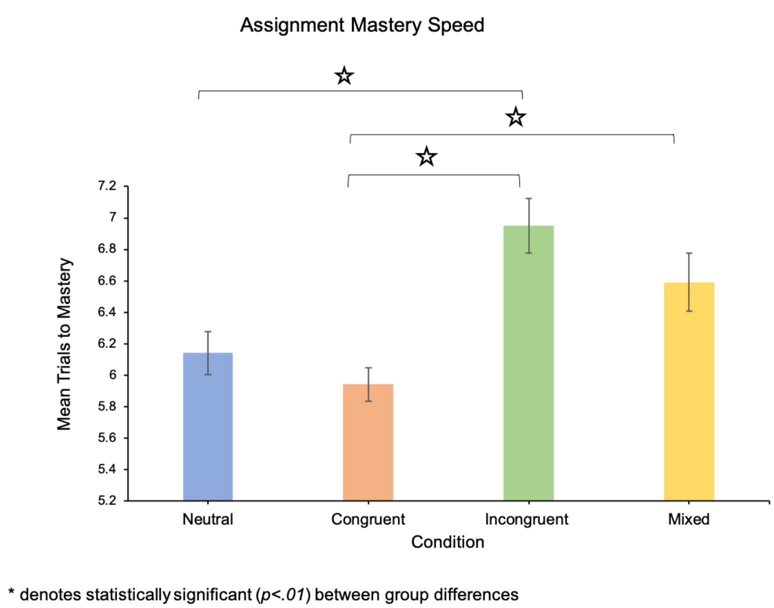
Population	*n*	Average Prior Performance (*SD*)	Average Mastery Speed (*SD*)
Overall	2,152	.70 (.14)	6.38 (3.24)
5th Grade	13	.86 (.15)	7.23 (5.00)
6th Grade	251	.71 (.13)	6.34 (2.96)
7th Grade	654	.67 (.14)	6.39 (3.13)
8th Grade	210	.76 (.15)	6.25 (3.72)
9th Grade	399	.71 (.13)	5.90 (2.44)
10th Grade	21	.77 (.11)	6.57 (3.16)
11th Grade	2	.72 (.09)	5.50 (.71)
12th Grade	24	.73 (.08)	6.79 (3.90)

Note: Grade level was not reported for *n*=578 participants.

Next, we conducted a preliminary one-way ANOVA to examine differences in average mastery speeds by condition. Results indicate that there were statistically significant overall differences between groups in mastery speed (F (3,2148) = 10.33, $p < 0.01$). Post hoc tests using Bonferroni correction to account for multiple comparisons revealed that, on average, students in the congruent condition (*M* = 5.94 problems, *SD* = 2.29 problems) mastered the assignment in significantly fewer problems than in the incongruent condition (*M* = 6.95 problems, *SD* = 3.58 problems, Cohen's *d* = 0.34) and the mixed condition (*M* = 6.59 problems, *SD* = 3.93 problems, Cohen's $d = 0.20$); see Figure 4. Students in the neutral condition ($M = 6.14$ problems, $SD = 2.91$ problems) also completed the problem set in significantly fewer problems than students in the incongruent condition (Cohen's $d = 0.25$). There were no differences in mastery speed between the mixed and incongruent condition ($p > 0.10$). These results prompted further exploration into examining the impact of spacing condition on assignment mastery speeds, accounting for grade level, prior performance, and the nesting of students in classrooms.

Figure 4

Mean Number of Trials Required for Mastery as a Function of Spacing Condition with Error Bars Reporting One Standard Error of the Mean on Each Side



Hierarchical Linear Models Examining the Impact of Spacing Condition on Mastery Speed

Table 3 (below) displays the results of the four two-level hierarchical linear models for these analyses. The unconditional model (Model 1) predicting mastery speed had an ICC of 0.097, indicating that 9.7% of the variance in assignment mastery speed is due to class level differences. The percentage of variance explained for each model is derived from the variance components of the model directly preceding it, explaining the variance accounted for above and beyond the previous model.

Model 2 shows the influence of students' grade level and prior performance in ASSISTments on assignment mastery speed. Prior performance was a significant predictor of mastery speed, where students with higher prior performance on ASSISTments problem sets had lower mastery speeds ($\beta = -2.31$, $p < 0.05$). Grade level did not significantly predict mastery speed ($p > 0.05$). The addition of these two variables explained 7% of the child level variance and 58% of the class level variance in mastery speed.

Model 3: Research Question 1: Does Spacing Influence Assignment Mastery Speed?

A 2-level HLM model (Model 3) was conducted to examine the impact of condition on mastery speed, controlling for grade and prior performance. The incongruent spacing condition was treated as the reference group for the hierarchical linear models since the ANOVA indicated that there were significant differences between the incongruent spacing condition and two other groups. Results were consistent with the ANOVA; there was a significant effect of two conditions on assignment mastery speeds. The analysis revealed that students in the congruent condition ($\beta = -0.92$, $p < 0.01$) and the neutral condition ($\beta = -0.78$, $p < 0.01$) mastered the assignment in significantly fewer problems than in the incongruent condition. Specifically, students in the congruent condition completed the assignment, on average, in .92 problems faster than students in the incongruent condition. Similarly, students in the neutral condition completed the assignment, on average, in .78 problems quicker than students in the incongruent condition. While both congruent and neutral spacing conditions significantly predict assignment mastery speed, the effect is larger for the congruent spacing condition than the neutral spacing condition.

Although there was a trend, there were no significant differences between the mixed condition and the incongruent condition (β = -0.46, $p < 0.07$). Higher prior performance was related to faster mastery speeds (β = -2.32, $p < 0.01$). Grade level was not significant in predicting mastery speed.

Table 3

Hierarchical Linear Models Show the Effect of Condition, Prior Performance, and Grade on Mastery Speed (Note: $^{+}$ *p <. 10, * p < .05, ** p < .01)*

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	6.42 (.11)**	6.23 (0.11)**	6.79 (0.22)**	7.67 (1.04)**	8.04 (1.09)**
Level-1 (Student)					
Grade Level		-0.01 (0.09)	-0.02 (0.09)	-0.02 (0.09)	-0.22 (0.20)
Prior Performance		-2.31 (0.69)**	-2.32 (0.70)**	-3.70 (1.61)*	-2.33 (0.70)**
Neutral			-0.78 (0.24)**	-2.21 (1.77)	-3.62 (1.72)*
Congruent			-0.92 (0.26)**	-2.68 (1.35)*	-2.43 (2.15)
Mixed			-0.46 (0.24) $^{+}$	-0.82 (1.70)	-2.33 (1.54)
Neutral x Prior Performance				2.02 (2.37)	
Congruent x Prior Performance				2.51 (1.77)	
Mixed x Prior Performance				0.52 (2.27)	
Neutral x Grade					0.37 (0.21)$^{+}$
Congruent x Grade					0.20 (0.25)
Mixed x Grade					0.25 (0.18)
Level-2 (Teacher)					
Variance Components					
Student Level	9.52	8.90	8.80	8.79	8.78
Teacher Level	1.02	0.43	0.42	0.43	0.43
Total Variance	10.54				
Level 1	0.903				
ICC Level 2	0.097				
% explained at student level		0.07	0.01	0.00	0.00
% explained at classroom level		0.58	0.02	-0.02	0.00

Model 4: Research Question 2: Does Prior Performance Moderate the Effect of Condition on Assignment Mastery Speed?

Next, we tested whether there was an interaction effect between students' prior performance and condition to understand if the effect of spacing condition was moderated by prior performance. Model 4 presents all Level-1 variables in addition to the prior performance interaction terms within Level-1. The interactions between prior performance and each spacing condition (neutral, congruent, mixed) were not significant predictors of assignment mastery speed (β = 2.02, $p = 0.39$; β = 2.51, $p = 0.16$; β = 0.52, $p = 0.82$), which means only the main effects of prior performance and spacing condition should be interpreted. This result suggests that students' prior performance in ASSISTments does not moderate the relationship between spacing condition and assignment mastery speed.

Model 5: Research Question 3: Does Grade Moderate the Effect of Condition on Assignment Mastery Speed?

Lastly, we tested whether there were interaction effects of grade level × condition on assignment mastery speed. Model 5 presents all Level-1 variables in addition to the grade level interaction terms within Level-1. The interaction between grade level and the neutral spacing condition was not significant but was trending towards significance (β = 0.37, $p = 0.08$). The other interactions between grade level and each spacing condition (congruent, mixed) were not significant predictors of assignment mastery speed (β = 0.20, $p = 0.43$; β = 0.25, $p = 0.18$). This finding suggests that students' grade level does not moderate the relationship between spacing condition and assignment mastery speed.

Discussion

The goal of this study was to explore whether manipulating the physical spacing between mathematical symbols would impact students' assignment mastery speed on order of operations problems. In addition to replicating the difficulty that algebra learners experience with incongruent spacing in order of operations problems, we were particularly interested in examining whether spacing effects exist in both younger and older students in authentic homework environments such as an online tutoring system. Two main findings emerged from this study: 1) students in the incongruent condition had slower mastery speeds (solving more problems to achieve mastery) than students in the congruent or neutral conditions, and 2) there were no significant interactions between grade level and condition, or prior performance and condition, on mastery speeds. Together, these results suggest that viewing incongruent spacing within mathematical expressions led to more errors and lower performance for most students, regardless of age or prior performance, compared to those who viewed problems with congruent or neutral spacing between terms.

Physical Spacing in Math Expressions Affects Student Performance

We predicted that viewing congruent or neutral spacing within problems would lead to faster mastery speeds compared to viewing problems with incongruent or mixed spacing. The results mostly supported this hypothesis; students who viewed problems with congruent or neutral spacing tended to master the assignment in significantly fewer problems than students who viewed problems with spacing that was incongruent with the order of precedence. However, there were no significant differences in mastery speed between the neutral and mixed condition.

One explanation for why congruent spacing may lead to greater performance over incongruent spacing is that visually modifying the physical spacing of terms may bias people to naturally group proximal terms into grouped objects (Wertheimer, 1950). Building on this visual spacing bias, one could argue that perceptually grouping terms to be congruent with the order of precedence could be more advantageous for students by providing perceptual cues that direct attention towards higher precedence operations in expressions, as if providing visual scaffolding for the order of precedence within expressions. If this is the case, then it could be hypothesized that students who saw the congruent grouping should also be more likely to have faster mastery speeds than students in the neutral spacing condition. However, there were no clear advantages of using congruent as opposed to neutral spacing. This finding was aligned with those of Landy and Goldstone (2010) who also found that operation error rates did not differ between congruent and neutral spacing conditions.

The finding that viewing congruent and neutral spacing led to higher performance than incongruent spacing is consistent with prior studies (e.g., Gómez et al., 2014; Jiang et al., 2013, 2014; Landy & Goldstone, 2010; Rivera & Garrigan, 2016). An interpretation of these results is that while the visual structure of mathematical notation creates perceptual groupings that cue interpretation and computation biases, this effect is stronger when those groupings are incongruent with mathematical rules, knowledge, and the order of precedence. It is possible that the difference in mastery speed by condition is due to a reliance on multiple perceptual cues that individuals use when solving order of operations problems. Further, these cues may work in a hierarchical structure where physical spacing acts as a first-order perceptual cue and operands act as second-order cues to interpret and act on mathematical notation. As a result, when presented with incongruent spacing, students may (incorrectly) attend to and rely more on perceptual groups when simplifying an expression than when presented with congruent or neutral spacing.

Other work has suggested that spacing is used as an action-guiding cue; incongruent spacing elicits errors while congruent and neutral spacing in mathematical notation helps facilitate improved performance. Consequently, viewing congruent spacing in expressions may not be significantly more helpful than viewing neutral spacing because the perceptual cues from physical spacing would be redundant to cues from operands. This notion is supported by findings that individuals attend to multiplication operands quicker than addition operands and treat narrow spacing between terms similarly to multiplication operands (Landy et al., 2008). For instance, in the expression “4+2*3”, the multiplication operand acts as a cue to group the “2” and “3” when the physical spacing is neutral. If the expression was presented with physical spacing congruent with the order of precedence, “4 + 2*3”, the physical spacing would only reinforce the grouping between the “2” and “3” but is not necessary. Conversely, physical spacing would be more of a perceptual cue if the expression was presented as “4+2 * 3”. Based on the results from this study and examples from the body of literature on this work demonstrating the influence of visual properties on performance with mathematics notation, perhaps the physical spacing within mathematical notation is a higher-order perceptual cue than operands which is why viewing incongruent spacing may be much more challenging for students.

Prior Math Performance, Grade Level, and Spacing Conditions

There is a common view that students’ computational errors are an indication of their conceptual misunderstandings about mathematics. Consistent with this idea, students’ prior performance in ASSISTments significantly predicted their mastery speed, suggesting that, on average, students with higher prior performance made fewer errors when solving order of operations problems. However, even when controlling for prior performance, incongruent spacing still affected student performance on the problem set. Additionally, while previous findings have largely focused on college-aged participants in laboratory settings (e.g., Gómez et al., 2014; Jiang et al., 2013; Landy & Goldstone, 2007a, 2007b, 2010), the current findings reveal that the effects of physical spacing also occur in younger students who are in the process of learning the rules of operation precedence and applying that knowledge in authentic homework settings. Regardless of age or prior performance in the tutoring system, viewing expressions with spacing that is incongruent with the order of precedence seems to be more difficult for students ranging from the 5th to 12th grade.

Taken together, these findings reinforce previous evidence that subtle changes in physical spacing can impact students' performance on computing order of operations problems regardless of the student's age or knowledge level. It seems that perceptual grouping, through spacing, may be acting as an irrelevant but substantial lure that is hard for students to ignore. More broadly, the differences in performance across conditions supports the notion that people use space as a perceptual cue when interpreting and acting on mathematical symbols. As such, these results provide further evidence that visual and perceptual processes can drive reasoning about mathematics computations (Landy & Goldstone, 2007a, 2007b) and highlight the conflicts between relevant (rule-ordered) and irrelevant (spacing) features in the presentation of mathematics.

Implications for Teaching and Learning

The current study suggests that minor and relatively meaningless changes to the visual presentation of mathematical notation have implications for how students interpret and use symbolic notation to perform computations. Although perceptual cues influence mathematical reasoning, few instructional approaches or interventions make use of the power of perception. Future learning interventions for algebra learners could include purposeful manipulations to the presentation of mathematical expressions and equations which could affect students' abilities to learn and apply arbitrary mathematical rules such as the order of precedence. More broadly, the prevalence of a spacing effect on mathematics performance across upper elementary through high school age groups poses interesting questions and may have theoretical implications about when perceptual cues begin to drive cognitive processes in learning and development. Future work could investigate how early spacing effects emerge in young learners and how spatial manipulations may drive students' cognitive processes and actions at younger ages.

Limitations

There are several limitations to the dataset as well as the methods used in these analyses. For instance, the Skill Builder structure (where students must correctly answer three problems in a row) is not a common approach used in classroom instruction and does not easily lend itself to a pretest/posttest design. The Skill Builder structure also allows students to stop working after they achieve mastery status in an assignment by correctly answering three problems in a row. As a result, participants only answered an average of six problems in the assignment. Additionally, since the final dataset excluded students who answered the first three problems correctly, this sample does not take into account the behavior of the highest-performing mathematics students on this particular problem set.

Another limitation of the study is that there was limited demographic data available on the students. Since ASSISTments problem sets were assigned by teachers around the country who use the platform for homework, we are unable to collect specific data about children, teachers, or the classroom context. While this is certainly a limitation, the fact that ASSISTments Skill Builders are used and assigned widely by teachers allowed for more ecological validity and a much larger sample size than would have been collected if this study was conducted by researchers with recruited teachers in local classrooms. Additionally, effects of spacing were found even while controlling for prior performance and grade level.

Future Directions

While the current project focused solely on assignment mastery, subtle spacing manipulations have been found to influence student problem-solving behavior at the action level (Jiang et al., 2013). Based on work such as that of Jiang and colleagues (2013, 2014), examining other aspects of behavior while students are problem-solving may be fruitful lines of research. Jiang and colleagues (2013) manipulated perceptual grouping in a similar study using the minus sign ("-") as the operand of interest then analyzed rates of "target errors" that would be the result of relying on perceptual grouping rather than arithmetic precedence (i.e. subtracting 14 from "20 - 4 + 10"). Jiang et al. (2013) found that physical spacing manipulations did, in fact, lead to higher rates of target errors. Investigating the kinds of mistakes that are made on incorrect problems could provide new insights about the role of spacing in expressions and how perceptual groupings drive student actions while problem-solving.

One difference in our study from prior work is that we included a mixed spacing condition. While findings suggest that the mixed spacing condition was more difficult for students compared to students in the congruent spacing condition, interestingly, there were no differences in mastery speed between the mixed condition and the neutral condition, or the mixed and incongruent spacing conditions. Given these mixed findings, future experimental work should further examine the effects of mixed spacing and test plausible mechanisms for the impact of perceptual spacing.

It is also important to note that, across conditions, students were provided with an optional hint on each problem to remind them of the rules of precedence. Roughly a third of participants viewed the hint at least once while working on the assignment. That said, little to no work has studied how conceptual knowledge reminders about the order of precedence may mitigate the effect of physical spacing on mathematics performance. To develop a richer understanding of how perceptual grouping may affect student behavior, our team is currently exploring patterns of student behavior within problems, such as response times, help-seeking behavior, and error types, to see how effects of spacing manipulations extend to a broader population of students. More thorough error analyses at the action level, within problems, could provide insight about whether students who are presented with incongruent spacing tend to make predictable errors based on how the symbols were visually grouped. We also plan to explore students' actions after viewing the available hint to examine whether order of precedence errors continue to occur after a visual reminder of the order of precedence.

Conclusion

This work demonstrates that irrelevant, but salient visual information in notation, such as spacing, can influence students' reasoning of mathematics. Specifically, perceptual cues, even those that are mathematically misleading such as incongruent spacing, are difficult to ignore. This study extends prior work in the following ways. First, to our knowledge, this is one of the first randomized controlled trials conducted on physical spacing in the context of an online learning platform with school aged children, showing that perceptual grouping occurs in authentic learning environments in addition to laboratory settings. Second, these results reveal that incongruent spacing between terms does impact 5th to 12th grade students' performance on mathematics assignments, resulting in slower mastery speeds, even when a reminder about the order of precedence is available as a hint on each problem. Conversely, terms in mathematical expressions that are neutrally spaced or grouped together to be congruent with the order of precedence increase assignment mastery speeds. These findings further support the notion that subtle perceptual cues, such as physical spacing, do not bear any practical implications in mathematics yet have effects on mathematical cognition and performance for students in upper elementary through high school. Learners and experts alike utilize perceptual strategies when reasoning about mathematics.

Funding

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1645629. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Additionally, we are grateful for the support of the Graduate Assistance in Areas of National Need fellowships #P200A150306 and #P200A180088.

Competing Interests

We have no competing interests to disclose.

Acknowledgements

We thank David Landy for his consultation on this project, as well as Janice Kooken and Anthony Botelho for their contributions to data analysis. We also thank the ASSISTments team for access to their research testbed.

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Botelho, A., Wan, H., & Heffernan, N. (2015, March). The prediction of student first response using prerequisite skills**.** In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, 39-45. ACM.

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.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. East Norwalk, CT: Appleton-Century-Crofts.

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.

Gómez, D. M., Benavides-Varela, S., Picciano, C., Semenza, C., & Dartnell, P. (2014). Symbol spacing and its influence in school children's calculation of mixed-operation arithmetic expressions. In Oesterle, S., Nicol, C., Liljedahl, P., & Allan, D. (Eds.). *Proceedings of the Joint Meeting of PME 38 and PME-NA 36*, Vol. 6, pp. 310. Vancouver, Canada: PME.

Gómez, D. M., Bossi, B., & Dartnell, P. (2014). Spacing patterns affect students' application of arithmetic precedence. In Oesterle, S., Nicol, C., Liljedahl, P., & Allan, D. (Eds.). *Proceedings of the Joint Meeting of PME 38 and PME-NA 36*, Vol. 6, pp. 310. Vancouver, Canada: PME.

Gunnarsson, R., Hernell, B., & Sönnerhed, W. W. (2012). Useless brackets in arithmetic expressions with mixed operations. In *PME36, the 36th Conference of the International*

Group for the Psychology of Mathematics Education, Taipei, July 18-22, 2012 (pp. 2-275). The International Group for the Psychology of Mathematics Education.

Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24, 470-497.

Hoch, M., & Dreyfus, T. (2004). Structure sense in high school algebra: the effect of brackets. In M.J. Høines & A.B. Fuglestad (Eds.), *Proc. 28th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 3, pp. 49-56). Bergen, Norway: PME.

Jiang, M. J., Cooper, J. L., & Alibali, M. W. (2013). Visually grouping operands: Perceptual factors influence arithmetic performance. In Martinez, M. & Castro Superfine, A. (Eds.). *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Chicago, IL: University of Illinois.

Jiang, M. J., Cooper, J. L., & Alibali, M. W. (2014). Spatial factors influence arithmetic performance: The case of the minus sign. *The Quarterly Journal of Experimental Psychology, 67*(8), 1626-1642.

Kirshner, D. (1989). The visual syntax of algebra. *Journal for Research in Mathematics Education, 20,* 274–287. doi: 10.2307/749516

Landy, D. H., Jones, M. N., & Goldstone, R. L. (2008). How the appearance of an operator affects its formal precedence. *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society*, pp.2109–2114. Washington, DC: Cognitive Science Society.

Landy, D., & Goldstone, R. L. (2007a). How abstract is symbolic thought? *J. Exp. Psychol. Learn. Mem. Cogn*. 33, 720–733. doi: 10.1037/0278-7393.33.4.720

Landy, D., & Goldstone, R. L. (2007b). Formal notations are diagrams: Evidence from a production task. *Memory & Cognition*, *35*(8), 2033-2040.

Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *The Quarterly Journal of Experimental Psychology*, *63*(10), 1953-1968.

Lee, V. E. (2010). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist, 35*(2), 125-141.

Linchevski, L., & Livneh, D. (1999). Structure sense: The relationship between algebraic and numerical contexts. *Educational Studies in Mathematics, 40*(2), 173-196.

Marchini, C. & Papadopoulos, I. (2011). Are useless brackets useful tools for teaching? In B. Ubus (Ed.), *Proc. 35th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 3, pp. 185-192). Ankara, Turkey: PME.

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), 25.

McNeil, N. M., & Alibali, M. W. (2004). You'll see what you mean: Students encode equations based on their knowledge of arithmetic. *Cognitive Science, 28,* 451-466.

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*, 883.

Niehaus, E., Campbell, C. M., & Inkelas, K. K. (2014). HLM behind the curtain: Unveiling decisions behind the use and interpretation of HLM in higher education research. *Research in Higher Education, 55*, 101-122.

Ostrow, K. S., Selent, D., Wang, Y., VanInwegen, E. G., Heffernan, N. T. & Williams, J. J. (2016). The assessment of learning infrastructure (ALI): The theory, practice and scalability of automated assessment. In the *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, pp. 279-288.

Rivera, J., & Garrigan, P. (2016). Persistent perceptual grouping effects in the evaluation of simple arithmetic expressions. *Memory & Cognition*, *44*(5), 750-761.

Rumelhart, D., Smolensky, P., McClelland, D., & Hinton, G. (1986). Schemata and sequential thought processes in pdp models. In *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 7–57). Cambridge: MIT Press.

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*, 1172–1217.

Walkington, C., Clinton, V., & Sparks, A. (2019). The effect of language modification of mathematics story problems on problem-solving in online homework. *Instructional Science*, 1-31.

Wertheimer, M. (1950). Gestalt theory. In W. D. Ellis (Ed.), *A sourcebook of Gestalt psychology* (pp. 1–11). New York, NY: Humanities Press.

Chapter 3. Resisting the Urge to Calculate: The Relation Between Inhibition Skills and Perceptual Cues in Arithmetic Performance

Chapter 3 presents an in-preparation manuscript that is being prepared in collaboration with Jenny Yun-Chen Chan and Erin Ottmar. This project was prompted by the hypothesis that students' inhibitory control may play a large role in how students perceive and react to perceptual cues embedded in math instructional materials. Whereas prior work has demonstrated general effects of perceptual cues on students' performance in math (e.g., Harrison et al., 2020), less is known about how different groups of students may be impacted differentially by viewing problems with perceptual cues like spatial proximity. A large body of research has demonstrated the power of inhibitory control on creative problem solving (e.g., Cassotti et al., 2016) and the same premise applies to findings in math and perceptual learning research. For instance, if primed, students are likely to attempt applying inappropriate procedural rules to a problem rather than first attempting any sense-making (Lawson et al., 2021). Similarly, in a project I have collaborated on (Chan et al., 2021), we found that students were more likely to solve problems on their first attempt when the problem was presented with numbers rather than symbols without varying in difficulty. Findings like these suggest that students may rely on applying procedural rules by rote rather than pausing to consider conceptual properties and rules that may influence their approach to problem solving. These findings could possibly be explained by a relation between students' behavior and inhibitory control as their ability to resist applying procedural rules blindly, motivating the study presented in this chapter.

Abstract

Subtle visual manipulations to the presentation of mathematical notation influence the way that students perceive and solve problems. While there is a robust impact of perceptual cues on students' problem-solving, other cognitive skills such as facets of executive functioning may interact with perceptual cues to impact students' arithmetic problem-solving performance. Currently, we present an online randomized controlled trial in which college students completed a version of the Stroop task followed by arithmetic problems presented with neutral, congruent, and incongruent spacing. We found that students were comparably accurate across problem types but spent longer responding to problems with congruent spacing. Further, inhibitory control was not a significant predictor of performance. These results advance perceptual learning theory and provide considerations for designing instructional materials by demonstrating how students are impacted by perceptual support in online settings.

Keywords: perceptual learning, inhibitory control, arithmetic

Resisting the Urge to Calculate: The Relation Between Inhibitory Control and Perceptual Cues in Arithmetic Performance

Students' mathematical skills and achievement are related to outcomes later in life such as their academic degree attainment and employment status (Adkins & Noyes, 2018; Parsons & Bynner, 2005). However, many students struggle to progress beyond basic algebra (Kena et al., 2015). This struggle may be partially due to the twofold challenge that students face: acquiring conceptual knowledge of mathematics principles and pairing that knowledge with a set of procedural rules to solve problems. Although problem-solving accuracy and efficiency are primary goals in mathematics education (Common Core State Standards, 2010), many students struggle to acquire structure sense — the ability to detect patterns and derive meaningful interpretations from mathematical notation in order to efficiently act on it (Hoch & Dreyfus, 2004; Livneh & Linchevski, 2007; Jupri et al., 2021). Therefore, instructional support for algebra should direct students' attention to important structural pieces of notation so that students perceive structural patterns that will help them manipulate equations efficiently.

One strategy to help students learn to notice and attend to important structures and patterns within mathematical notation is to leverage perceptual scaffolds. Perceptual scaffolds are subtle visual features of instructional materials that intentionally cue students' attention towards pieces of information to aid with accurate and efficient problem-solving (e.g., strategic coloring, spacing, sizing, and positioning of symbols; Closser et al., under review; Goldstone et al., 2010, 2017). In particular, students simplify expressions with greater accuracy when viewing problems with less space between higher-order operands, demonstrating that perceptual features impact mathematical reasoning when solving problems (Harrison et al., 2020). Alongside exposure to perceptual support during problem-solving practice, extensive evidence has shown that individual differences in executive function skills (i.e., brain functions that control working memory, flexibility, and inhibition) are related to, and predictors of, creative problem solving (Cassotti et al., 2016) and performance in mathematics (Cragg & Gilmore, 2014). We propose to extend this research by investigating the role that inhibitory control plays in learning from instructional materials that contain perceptual cues.

In this study, we investigate how subtle variations in perceptual features impact students' problem-solving performance and how that effect may vary for students with different levels of inhibitory control. College students participated in an online experiment where they completed a version of the Stroop task (to measure inhibitory control) followed by blocks of order-of-operation problems that presented symbols with spacings congruent, neutral, and incongruent to the order of operations, respectively. Through this study, we aim to conceptually replicate and extend prior work on the effects of perceptual cues in online mathematics materials by: 1) testing for an effect of spacing on students' problem-solving accuracy and response time, and 2) investigating whether and how that effect may be moderated by individual differences in students' inhibitory control.

Perceptual Learning: Theory and Mechanisms

While seemingly abstract, math—and the ways we learn and reason about it—are influenced by our perceptual processes (Goldstone et al., 2010; Kellman et al., 2010; Marghetis et al., 2016). According to theories of perceptual learning, students change how they perceive incoming information over time, directing their attention towards more relevant information in their environment to support high-level cognition (Gibson, 1969; Goldstone, 1998). Visual features of instructional materials can influence this process to impact students' mathematical reasoning and learning.

Although perceptual features do not alter the mathematical meaning of notation, subtle visual manipulations to instructional materials do impact the way that students both *process* incoming information and *act* on it (Jiang et al., 2014). Rivera and Garrigan (2016) found that perceptual grouping effects impact the way that undergraduate students process expressions; even verbalizing arithmetic expressions after briefly viewing them does not negate the effect of perceptual grouping cues. By changing how students process mathematical notation, perceptual features also impact how students act on notation to perform transformations (Goldstone et al., 2010). For example, uniquely coloring important pieces of notation can be used as a perceptual grouping mechanism to connect different representations, improve students' ability to generate problem-solving strategies, and support learning (Alibali et al., 2018; Chan et al., 2019).

Other common perceptual cues in mathematics education research seem to leverage Gestalt principles of grouping by creating a common visual region so that individuals perceive a group rather than individual objects (Landy & Goldstone, 2007; Wagemans, 2012). For instance, middle-school students are more likely to combine pairs of numbers that match the problem-solving goal when the numbers are adjacent (e.g., transform 5 + 33 + 7 into 5 + 40) vs. non-adjacent to each other (e.g., transform 33 + 5 + 7 into 5 + 40; Lee et al., 2022). A more explicit example is the use of superfluous parentheses: parentheses which create a perceptual group of terms without altering the mathematical meaning of notation (e.g., 7+(5*6)−2). Viewing superfluous parentheses has increased students' problem-solving performance and their understanding of notational structure (Hoch & Dreyfus, 2004; Papadopoulos & Gunnarsson, 2020). These examples demonstrate how perceptual cues can influence the ways that students interpret incoming information and impact students' mathematics performance in multiple ways. Further, research has shown that even experts tend to rely on perceptual cues when solving complex equations, suggesting that reliance on perceptual cues may increase with proficiency as well as with increased cognitive demands to solve complex problems (Braithwaite et al., 2016; Rumelhart et al., 1986).

Spacing as a Perceptual Grouping Mechanism

Here, we focus on physical spacing around operands as a perceptual grouping mechanism. A large body of research has demonstrated that students are impacted by the physical distance between terms and operands in mathematics expressions; specifically, when terms and operands are closer together they are viewed as a group (e.g., 6+8) whereas when they are spaced farther apart they are viewed as individual objects (e.g., 6 + 8). In order-of-operations problems with multiple operators, physical spacing can be leveraged to draw students' attention towards the perceived groups of objects and influence their problem-solving performance. Specifically, students demonstrate higher problem-solving performance when they view problems with spacing that is congruent with the order of operations (e.g., 3 + 4×2) and lower performance when viewing problems that are presented with spacing that is incongruent with the order of operations (e.g., 3+4 × 2; Braithwaite et al., 2016; Gomez et al., 2014; Harrison et al., 2020; Landy & Goldstone, 2010). This phenomenon persists across lab settings with undergraduates (e.g., Landy & Goldstone, 2010) as well as in authentic learning environments with K-12 students (e.g., Braithwaite et al., 2016; Gomez et al., 2014). Previously, our team also extended this line of research by showing that regardless of grade and prior knowledge, secondary students who viewed expressions with neutral (e.g., 3+4×2) or congruent spacing were more accurate on solving arithmetic problems in an online homework assignment than their peers who viewed expressions with incongruent spacing (Harrison et al., 2020). These findings demonstrate that spacing can be a powerful form of perceptual scaffolding that is invariably difficult to ignore.

In the current study, we first aim to conceptually replicate previous findings on perceptual learning by testing if and how spacing impacts college students as they solve order-of-operations problems. Second, we plan to compare differences in students' problem-solving accuracy as well as response times to consider how perceptual cues impact both performance and efficiency. Landy and Goldstone (2010) did compare students' response times on problems that differed in the physical spacing of the symbols but they did not find significant differences between these problems. Their results motivate our third goal: to move beyond detecting general effects of spacing by investigating how individual differences in inhibitory control impacts the effect of spacing on students' problem-solving performance.

Inhibition and Mathematics

Inhibition is defined as the skill to suppress habitual responses or ignore distracting information. It is an important skill that enables students to ignore irrelevant facts or suppress a habitual procedure (Gilmore et al., 2015; Lee & Lee, 2019). Cragg and Gilmore (2014) posit that inhibition may play an important role early in mathematics education as children learn to suppress less sophisticated problem-solving strategies (e.g., counting all the numbers) in order to adapt more sophisticated strategies (e.g., counting from the larger addend). Students may also use inhibition as they learn about related but different number facts and concepts (e.g., inhibiting $2 + 3 = 5$ when asked $2 \times 3 =$ __). Further, Lee and Lee (2019) suggest that students may call on their inhibitory control to: (a) ignore distractions in the classroom, textbook, or problems, (b) suppress misconceptions (e.g., "=" means total) while prioritizing correct but not well-learned information (e.g., "=" means both sides have the same value), and (c) inhibit well-learned procedures (e.g., whole-number addition) while drawing on appropriate knowledge (e.g., fraction operations).

Despite the conceptual connections between inhibition and mathematical performance, the findings on this association are inconsistent (see Bull & Lee, 2014) and have led to calls for understanding how multiple cognitive mechanisms interact to impact students' mathematics skills (Wilkey et al., 2019). Part of the issue may be that inhibition skill is often measured with laboratory tasks (e.g., stop signal task, day-night task, head toes knees and shoulders task, Stroop task) that may not accurately represent or reflect how inhibitory control is involved in mathematics problem-solving (Lee & Lee, 2019). Further, depending on the types of stimuli used in tasks (e.g., numerical vs. non-numerical), the relation between inhibitory control and mathematics performance also varies (Bull & Scerif, 2001; Wilkey, et al., 2019), suggesting that this relation may depend on the types of tasks as well as the stimuli within the task used to measure inhibitory control.

Findings on various versions of the Stroop task provide mixed insights into the association between aspects of students' inhibitory control and their mathematics performance. For instance, Bull and Scerif (2001) found that inhibition skill, as measured by the number-quantity Stroop task (i.e., participants were instructed to say the name or the number of numerals such as "222"), was correlated with six- to eight-year-olds' mathematics performance; however, this association was not significant when children's inhibition skill was measured by the color-word Stroop task (i.e., participants were instructed to say the color of the words instead of reading the word). Further, Kroesbergen and colleagues (2009; JPEA) found that five- to seven-year-olds' inhibition skill, as measured by the animal-size Stroop task (i.e., select the animal that is larger in real life when the animal images vary in size on-screen), predicts their early mathematical competence above and beyond language, updating, and planning skills. However, Bellon and colleagues (2016) found that eight- and nine-year olds' performance on number-quantity Stroop task and the color-word Stroop task did not correlate with or predict their mathematical performance on arithmetic fact retrieval. These findings demonstrate that while inhibitory control may be related to mathematical performance, this relation may depend on the types of inhibition tasks, the stimuli used in these tasks, and their alignment with the mathematical skill of interest.

In the current study, we examine the relation between students' performance on order-of-operation problems and their inhibitory control as measured by the animal-size Stroop task. We chose the animal-size Stroop task because it is (a) a non-numerical measure and (b) related to perception of spatial relations. Specifically, if we find a significant association between students' performance on the animal-size Stroop task and the order of operation problems, the finding may reflect a genuine association as we cannot attribute this association to the numerical stimuli. Further, the animal-size Stroop task involves perceiving the spatial relations between stimuli, aligning with the perception of spacing in the order of operation problems.

Current Study

To extend prior research on perceptual grouping mechanisms and further investigate the potential relations between inhibition and mathematics performance, we conducted an online within-subjects experiment. College students at a private university completed a version of the animal-size Stroop task, as a measure of their inhibitory control, followed by a series of order-of-operations problems presented with neutral, congruent, and incongruent spacing. Our research questions are as follows:

1. *How does students' performance (i.e., accuracy, response time) on order-of-operation problems vary between problems with congruent (e.g., 5*4 + 3) and incongruent (e.g., 5 * 4+3) spacing, controlling for performance on neutrally spaced (e.g., 5*4+3) problems?* Based on prior research, we hypothesize that students would be more accurate and quicker on the problems with congruent spacing as opposed to incongruent spacing.
2. *How do students' problem-solving performance (i.e., accuracy, response time) vary by their inhibitory control?* Based on prior research, we hypothesize that students with higher inhibitory control would be more accurate and quicker on all of the order-of-operation problems.
3. *Is there an interaction between students' inhibitory control and performance on order-of-operation problems presented with congruent vs. incongruent spacing?* We plan to explore this relation and do not have an a priori directional hypothesis.

Methods

The plan for this study was pre-registered on Open Science Framework prior to data collection (https://osf.io/bh6kx) and received approval from our university's Institutional Review Board.

Participants

An a priori power analysis conducted in G*Power determined that a sample size of 186 students would provide sufficient power (.90) and confidence (p <.05) to detect a small to medium effect size of f = .12, comparable to the size of perceptual effect reported in prior research (Harrison et al., 2020). To meet the sample size, we recruited a total of 233 students taking undergraduate courses at a private university in the northeastern U.S. through the online participant pool for psychology students. Students were compensated for their time with partial course credit. Of the 233 students who started the experiment, 196 completed the Stroop task, baseline items, and at least multiple problems in both experimental conditions. Of those students, six students were excluded due to a data logging error and 16 were excluded for outlier performance, leaving 174 students included in the analytic sample. A post hoc power analysis in G*Power determined that this analytic sample size would still allow us to detect a small to medium effect size of $f = .12$ with sufficient power (.88) and confidence (p <.05).

Of the 174 students, we received information on age from 172 students (17-27 years old; $M = 19.48$ years, $SD = 1.50$ years) and year in school from 171 students. Of the 171 students, 57 (33%) students reported being in their first year at the university, 42 (25%) in their second year, 34 (20%) in their third year, 32 (19%) in their fourth year, and two (1%) students in their fifth year. Additionally, one student reported being in high school and three students reported "other". We also received self-reported gender information from 170 students: 95 (56%) females, 67 (39%) males, seven (4%) non-binary, and one agender participant.

Study Design and Procedure

Participating students clicked a URL link to complete the online study in a web browser on their personal devices. As students opened the study, they first completed a version of the Stroop task designed to assess inhibition skill levels. Following the Stroop task, students completed three blocks of 16 order-of-operations problems. First, students completed problems presented with neutral spacing (control; e.g., 4 × 3 − 10). Next, students completed problems presented with congruent spacing (experimental; e.g., 4×3   − 10) or incongruent spacing (experimental; e.g., 4 × 3−10). We used a within-subjects design that randomly counterbalanced the order of experimental conditions between students; specifically, students either completed the congruent problems followed by incongruent problems or vice versa. Finally, students were asked to report their age, gender, and race at the end of the session.

All tasks were programmed using Psychophy, an open-source software for behavioral experiments, and administered through Pavlovia, an online platform for behavioral data collection.

Materials

Order-of-Operations Problems

We systematically created three blocks of 16 order-of-operations problems. Each problem included two operands: multiplication and either addition or subtraction. On half of the problems within each block, the multiplication operand was positioned on the left side of the expression, and addition (e.g., 3×5+7; four problems) or subtraction (e.g., 6×2−8; four problems) was positioned on the right side of the expression. On the other half of the problems, multiplication was positioned on the right side of the expression, and addition (e.g., 3+5×7; four problems) or subtraction (e.g., 6−2×8; four problems) on the left side of the expression. The numbers in each problem include one small (1, 2, or 3), medium (4, 5, or 6) and large (7, 8, or 9) value that were systematically varied in their position from left to right (e.g., 2×4+7). The correct answer on all problems were integers ranging from −53 to 50.

Across the three blocks, the problems followed the same rules and matched on the structure and magnitude. However, none of the problems were identical so students would not be able to recall an answer from a previous problem within the study. Across the three blocks, the spacing between the numbers and the operators varied between neutral (e.g., 4 × 3 − 10), congruent (e.g., 4×3 - 10), or incongruent (4 × 3−10) to the order of operations. The neutral spacing block served as a measure of students' prior knowledge, allowing us to examine the influence of students' inhibitory control, independent of their prior knowledge, on their performance on the congruent and incongruent blocks.

Stroop Task

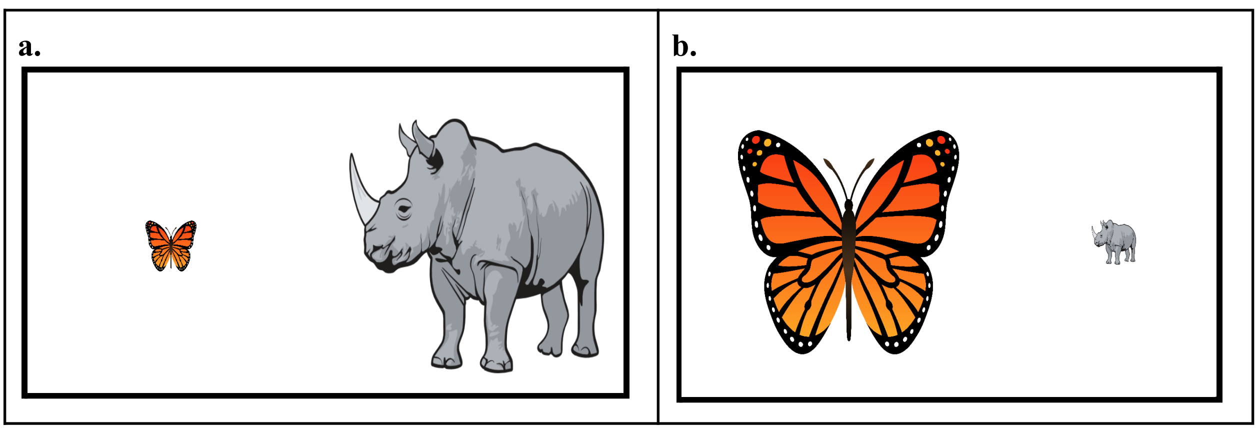
Students completed a version of the Stroop task created based on Szucs et al. (2013) to assess their levels of inhibitory control. The Stroop task was administered with four practice trials followed by 64 test trials.

First, students completed four practice trials without receiving any feedback. Students received the following instructions: "*In this task, you will see some animal pictures on two sides of the screen. Use your mouse to select the animal that is bigger in real life.*". On each practice trial, students first saw a fixation "+" in the middle of the screen for 750 milliseconds followed by two animal images on the screen. Students saw a small animal (i.e., butterfly) on one side of the screen and a big animal (i.e., rhinoceros) on the other side. Since students were instructed to select the animal that is larger in real life, the rhino was the correct answer in all four of the practice trials (e.g., both Figure 1a and 1b). On each trial, the image size of the two animals varied systematically so that one of the images was five times bigger than the other. On half of the trials, the image sizes were congruent with the animal sizes; on the other half of the trials, the image sizes were incongruent with the animal sizes. Specifically, on the congruent trials, the rhinoceros was presented as a bigger image on screen (Figure 1a); on the incongruent trials, the rhinoceros was presented as a smaller image on screen (Figure 1b). The location (left or right) in which the correct answer appeared was counterbalanced across trials.

After four practice trials, students received a reminder to "*select the animal that is bigger in real life*" as quickly and accurately as possible, then proceeded to complete 64 test trials. On each test trial, students first saw the fixation cross in the middle of the screen for 500 milliseconds followed by an image of a small animal (i.e., mouse, frog, bird, or rabbit) on one side of the screen and an image of a big animal (i.e., elephant, horse, cow, or lion) on the other side. Similar to the practice trials, the image sizes were congruent with the animal sizes on half of the test trials and incongruent on the other half of the test trials. All students saw the same 64 trials: the location of the correct answer was counterbalanced across trials, and the order in which the students received each trial was randomized. Based on prior literature, the congruent trials should be easier for students because the animal that is larger in real life is also larger on-screen; the incongruent trials require students to exercise inhibitory control to ignore the image sizes on-screen and respond based on animal sizes in real life (Szucs et al., 2013).

Figure 1

Sample Stroop Task Practice Trials



Note: students must select the animal that is larger in real life. In trial a, the image size is congruent with the animal size in real life. In trial b, the image size is incongruent with the animal size in real life.

Measures

Performance by Problem Type: Congruent and Incongruent Problems. Students' performance on order-of-operation problems in the congruent and incongruent spacing blocks was used as the dependent measures in our primary analyses. Performance on the congruent problems was measured by a) accuracy and b) median response time on correctly answered congruent problems (Landy & Goldstone, 2010). Accuracy represented the percentage of problems answered correctly out of the 16 congruently-spaced problems that students completed. Response time was measured in seconds from the moment that each congruent problem appeared on-screen until the student entered an answer and clicked "next" on the screen. Similarly, we measured students' accuracy on incongruent problems and median response time on their correct responses, and used these two measures as indicators of their performance on incongruent problems.

Baseline Performance. Students' a) accuracy and b) median response time on correctly answered neutrally-spaced problems were used as a proxy measure of their baseline performance on order-of-operations problems. Baseline accuracy represented the percentage of problems answered correctly out of the 16 neutrally-spaced problems that students completed. Baseline response time was measured in seconds from the moment that each problem appeared on-screen until the student selected an answer and clicked "next" on the screen. The average accuracy on the neutrally-spaced problems was included as a covariate in the analyses comparing students' accuracy on congruent vs. incongruent problems. Similarly, the median response time on correctly responded neutral problems was included as a covariate in the analyses on response time.

Inhibitory Control. Students' performance on the Stroop task was used as a measure of their inhibitory control. Per Gilmore and colleagues' computation for the inhibition score (2015), we identified the median response time for correctly answered congruent and incongruent trials, then took the difference in the median response time on the two types of trials (i.e., $RT_{incongruent}$ –

$RT_{congruent}$). This difference score was used as an indicator of students' inhibition skill; specifically, a larger difference indicated a lower level of inhibitory control.

Approach to Analysis

Students who completed the Stroop task as well as all order-of-operations problems were included in analyses. Prior to primary analyses, we conducted descriptive and correlation analyses to examine the distribution of, and relations between, each measure as well as to inform the primary analyses. A total of 15 students demonstrated outlier performance (i.e., three standard deviations above or below the mean) on one or more outcome variables (i.e., accuracy on congruent- and incongruent-spaced problems, response time on congruent- and incongruent-spaced problems), and were excluded from the analytic sample. Additionally, one student was excluded from the analytic sample due to outlier performance on the Stroop task.

To answer our research questions, we conducted two repeated measures ANCOVAs using JASP software (JASP Team, 2020). The ANCOVAs treated condition (i.e., incongruent spacing, congruent spacing) as a within-subjects independent variable, inhibitory control as a continuous between-subjects independent variable, and performance on neutral-spaced problems as a covariate (i.e., accuracy or average response time, respectively). Specifically, we conducted an ANCOVA comparing students' average accuracy between congruent and incongruent problems and an ANCOVA comparing students' median response time between congruent and incongruent problems. First, we tested the main effect of condition in both ANCOVAs to answer whether students' performance on problems varies between viewing problems with congruent and incongruent spacing. Second, we tested the main effect of inhibition in both ANCOVAs to demonstrate whether students' performance on order-of-operations problems varies between students with higher or lower inhibitory control. Third, we examined whether an interaction existed between students' inhibitory control and performance in each ANCOVA to observe whether students' inhibitory control seems to moderate the effect of spacing in order-of-operations problems on students' performance.

We also used JASP software to conduct these analyses using Bayesian statistics in order to provide both frequentist and Bayesian interpretations of the results. Whereas a frequentist analysis of variance tests an alternative hypothesis, a Bayesian analysis of variance also directly tests the null hypothesis. Here, we used the default, non-informative prior specifications in JASP as recommended (Faulkenberry et al., 2020; van de Schoot & Depaoli, 2014). The default specification uses a JZS (multivariate Cauchy) prior on the effect scales with a default scale of 0.5. We interpreted the Bayes Factor (BF_{10}) based on common thresholds (Schönbrodt & Wagenmakers, 2018). Specifically, a BF_{10} value smaller than 1, 1/3, or 1/10 respectively provides anecdotal, moderate, or strong evidence for the null hypothesis. Similarly, a BF_{10} value greater than 1, 3, or 10 respectively provides anecdotal, moderate, or strong evidence for the alternative hypothesis.

After conducting our pre-registered analyses, we conducted exploratory analyses in order to better explain the reported results. First, we used a median split to divide participants into higher (i.e., perfect) and lower baseline accuracy groups then conducted 2 (Problem Type: Congruent or Incongruent) x 2 (Baseline Accuracy: Higher or Lower) repeated measures ANOVAs to explore how students with different degrees of prior knowledge may be differentially impacted by the perceptual cues on a) accuracy and b) response time on order-of-operations problems. Next, we used a median split to group students with higher and lower levels of inhibitory control. We conducted 2 (Problem Type: Congruent or Incongruent) x 2 (Inhibitory Control: Higher or Lower) repeated measures ANOVAs to explore how students with different degrees of inhibitory control may be differentially impacted by the perceptual cues on a) accuracy and b) response time on order-of-operations problems within our high-performing sample.

Results

Preliminary Analysis

Tables 1 and 2 respectively present the descriptive statistics of, and correlations, between each of our focal variables. As seen in Table 1, undergraduate students displayed high performance across baseline (neutral-spaced problems), congruent, and incongruent order-of-operations problems, confirming that they were able to solve these order of operations problems quickly and accurately. As expected, students demonstrated very high accuracy on the overall Animal Stroop task (*M* = .99, *SD* = .01) as well as on the congruent (*M* = .998, *SD* = .009) and incongruent tasks respectively (*M* = .91, *SD* = .02); therefore, we used a previously established computation of performance by taking the difference of students' median response times on each item type as our measure of inhibitory control (Gilmore et al., 2015). As seen in Table 1, this measure of inhibitory control was positive, suggesting that students, on average, were slower at making a correct response when the animal size was incongruent vs. congruent to the image size.

Table 1

Descriptive statistics of each variable (N = 174)

	Mean	SD	Min–Max	Skewness	Kurtosis
Inhibitory Control (s)	0.18	0.10	-0.15-.47	-0.08	0.22
Baseline Accuracy	0.94	0.09	0.5-1.0	-2.44	7.50
Congruent Problems: Accuracy	0.94	0.08	0.62-1.0	-1.74	3.43

Incongruent Problems: Accuracy	0.94	0.07	0.69-1	-1.4	2.13
Baseline Response Time (s)	5.23	1.35	2.54-11.47	1.07	2.54
Congruent Problems: Response Time (s)	5.63	1.59	2.59-9.97	0.52	-0.11
Incongruent Problems: Response Time (s)	5.40	1.48	2.42-10.8	0.57	0.33

Note. Abbreviation: (s) = seconds, SD = standard deviation, Min = minimum, Max = maximum.
$^{*}p<.05$. $^{**}p<.01$. $^{***}p<.001$.

Two findings were worth noting in the correlation analyses (Table 2). First, as expected, students' accuracy on the three types of problems were positively correlated, suggesting the potential need to control for students' baseline performance when testing the unique effects of spacing and inhibitory control. Second, students with higher baseline accuracy also tended to make correct responses faster, suggesting that accuracy and response time may capture related but distinct aspects of performance, supporting our plan to examine both in this paper.

Table 2

Correlations Between Each Variable

	Inhibitory Control	Baseline Accuracy	Congruent Problems: Accuracy	Incongruent Problems: Accuracy	Baseline Response Time (s)	Congruent Problems: Response Time (s)	Incongruent Problems: Response Time (s)
Inhibitory Control	–						
Baseline Accuracy	-.075	–					
Congruent Problems: Accuracy	-.071	.339***	–				
Incongruent Problems: Accuracy	-.019	.260***	.360***	–			
Baseline Response Time (s)	.144	-.174*	-.228**	-.124	–		
Congruent Problems: Response Time (s)	.143	-.186*	-.084	-.149	.696**	–	
Incongruent Problems: Response Time (s)	.080	-.186*	-.156*	-.134	.619***	.820***	–

Notes: Values represent Pearson's r for the correlation coefficient. Abbreviation: (s) = seconds.
*$p<.05$. **$p<.01$. ***$p<.001$.

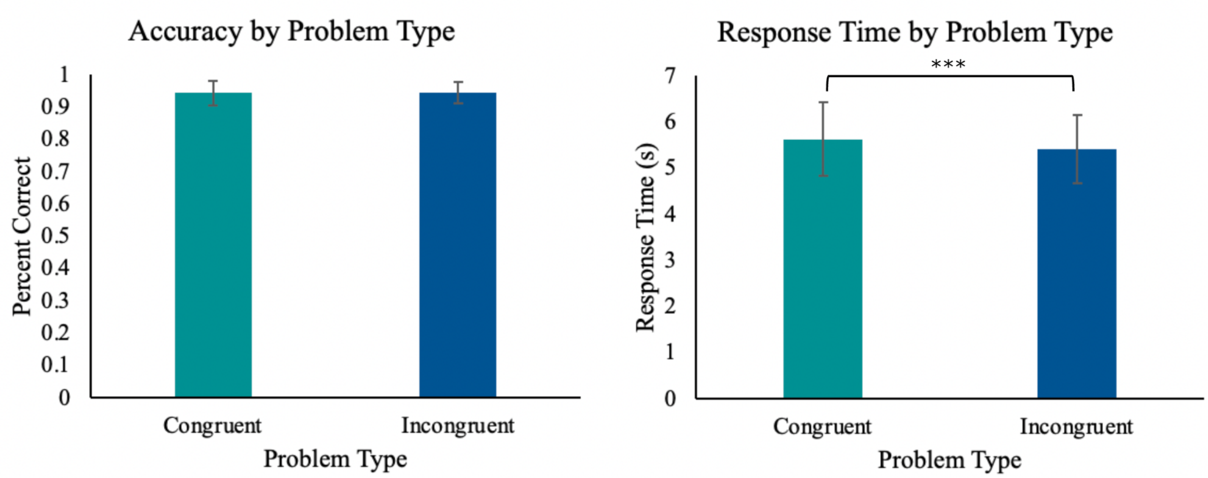
Effects of Physical Spacing on Congruent and Incongruent Problem Solving

Pre-registered Analysis: Effect of Perceptual Cues Controlling for Baseline Accuracy

To answer our first research question regarding the effects of problem type (i.e., congruent vs. incongruent spacing) on problem-solving performance, we conducted two sets of repeated measures ANCOVAs on students' accuracy and response time, respectively (Figure 2).

Figure 2

Left: *Accuracy by Problem Type. Right: Response Time By Problem Type in Seconds.*



Note: Error bars represent the standard deviation.
*** p = .001

A repeated measures ANCOVA controlling for students' baseline accuracy showed that the effect of problem type on problem-solving accuracy was not significant, (F[1, 172] = 1.862, $p = 0.174$, $\eta^2 = 0.003$), suggesting that students were comparably accurate on congruent and incongruent problems (Figure 2: Left). Students' baseline accuracy was a significant covariate, indicating that higher baseline accuracy was associated with higher accuracy on congruent and incongruent problems (F[1, 172] = 26.577, $p < .001$). A Bayesian ANCOVA confirmed the results with moderate evidence that problem type had no effect on problem-solving accuracy (BF_{10} = 0.118) and strong evidence that baseline accuracy had an effect (BF_{10} = 15685.816).

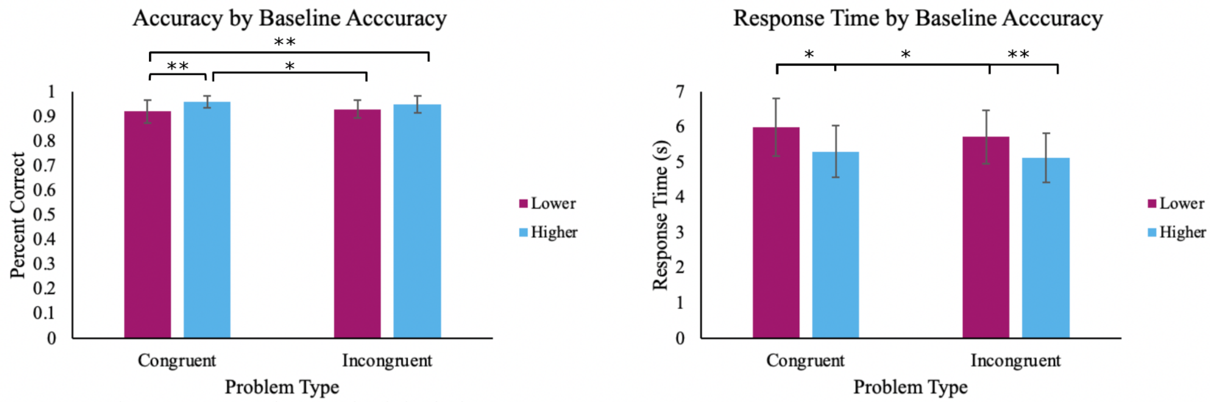
A repeated measures ANCOVA on students' median response time revealed a marginal difference between the two problem types ($F[1,172] = 3.308$, $p = 0.071$, $\eta^2 = 0.002$). Students took longer to solve congruent problems (*M* = 5.63 seconds, *SD* = 1.59) than incongruent problems (*M* = 5.40 seconds, *SD* = 1.48; Figure 2: Right). The Bayesian ANCOVA provided strong evidence in favor of the effect of problem type (BF_{10} = 13.090). There was also a significant effect of baseline response time ($F[1, 172] = 156.615$, $p < .001$), with strong evidence supporting this finding ($BF_{10} = 4.309*10^{22}$). Further, there was a Problem Type × Baseline Response Time interaction ($F[1,172] = 7.348$, $p = 0.007$, $\eta^2 = 0.004$), with strong evidence supporting this finding ($BF_{10} = 5.442*10^{23}$).

Exploratory Analysis: Effect of Perceptual Cues by Baseline Performance

The pre-registered analyses demonstrate that this sample is high-performing without much variance in students' accuracy on the order-of-operations problems. Further, the pre-registered analyses both showed that students' baseline accuracy was a significant predictor of accuracy and response. To delve further into how baseline accuracy might impact students' reactions to perceptual cues in order-of-operations problems, we conducted a median split on baseline accuracy (*Mdn* = 1.00) to compare students with lower accuracy (*n*=82) and higher accuracy (*n* = 92). This approach allowed us to explore how students with different levels of expertise on order-of-operations problems are impacted by perceptual cues on both accuracy and response time.

Figure 4

Left: Accuracy by Baseline Accuracy. Right: Response Time by Baseline Accuracy.



Note: Error bars represent the standard deviation.
* $p < .05$ ** $p < .01$

First, a 2 (Problem Type) × 2 (Baseline Accuracy: higher- vs. lower-performing) repeated measures ANOVA on students' problem-solving accuracy revealed that: a) students were comparably accurate on congruent vs. incongruent problems ($F[1,172] = 0.138$, $p = 0.710$, $\eta^2 = 2.532 * 10^{-4}$), and b) students with higher vs. lower baseline accuracy were more accurate on both types of problems ($F[1,172] = 14.226$, $p < 0.001$, $\eta^2 = 0.052$). Further, the Problem Type × Baseline Accuracy interaction was marginally significant ($F[1,172] = 3.714$, $p = 0.056$, $\eta^2 = 0.007$; Figure 4: Left). Post hoc comparisons with Bonferroni corrections showed that students with lower baseline accuracy were significantly less accurate on congruent problems ($M = .92$, $SD = .09$) than students with higher baseline accuracy on both congruent ($M = .96$, $SD = .05$) and incongruent problems $M = .95$, $SD = .07$; ps < .01). Additionally, higher-accuracy students were significantly more accurate on congruent problems than lower-accuracy students were on incongruent problems ($M = .93$, $SD = .07$; $p = .027$). A repeated measures Bayesian ANOVA revealed moderate evidence in support of the null effect of problem type ($BF_{10} = 0.118$), strong evidence in support of the effect of baseline accuracy ($BF_{10} = 97.568$), and the model including the Problem Type × Baseline Accuracy interaction ($BF_{10} = 11.239$).

Next, a 2 (Problem Type) × 2 (Baseline Accuracy) repeated measures ANOVA on students' response times revealed that: a) students were significantly slower to answer congruent problems than incongruent problems (F[1,172] = 10.308, p = 0.002, η^2 = 0.005), and b) there was a significant effect of baseline accuracy on response time (F[1,172] = 8.787, p = 0.003; Figure 4: Right). There was no significant Problem Type × Baseline Accuracy interaction (F[1,172] = 0.634, p = 0.427). Post hoc comparisons with Bonferroni corrections confirmed that students' spent longer on congruent (M = 5.63, SD = 1.59) than incongruent problems (M = 5.40, SD = 1.48; p = .002). Further, students with higher baseline accuracy demonstrated quicker response times on both problem types than students with lower baseline accuracy (p = .003). Further, students who with lower baseline accuracy had significantly longer response times on congruent (M = 6.00, SD = 1.64) than incongruent problems (M = 5.72, SD = 1.52; p = .039) and significantly longer response times than students with higher baseline accuracy on both congruent (M = 5.30, SD = 1.47; p = .015) and incongruent trials (M = 5.13, SD = 1.40; p = .001). A repeated measures Bayesian ANOVA revealed strong evidence in support of the effect of problem type (BF_{10} = 12.668) and strong evidence in support of the effect of baseline accuracy (BF_{10} = 10.139).

Effects of Inhibitory Control on Order of Operations Problems

Pre-registered Analysis: Effects of Inhibitory Control with the Baseline Covariate

To answer our second research question, we added students' inhibitory control performance as a continuous variable to the ANCOVA models. Aligned with the previous results, there was no effect of problem type (F[1,171] = 1.412, p = 0.236), but the effect of baseline accuracy was significant (F[1,171] = 25.988, p < .001) with strong evidence (BF_{10} = 19227.199). The effect of inhibitory control on students' problem-solving accuracy was not significant (F[1, 171] = 0.165, p = 0.685), with moderate evidence supporting this null finding (BF_{10} = 0.214). There was no interaction between problem type and inhibitory control (F[1, 171] = 0.313, p = 0.576).[1]

We repeated the analyses with response time, and still found the significant effect of problem type such that students had slower responses on congruent than incongruent problems (F[1,171] = 4.386, p = 0.038, η^2 = 0.002) with strong evidence supporting this finding (BF_{10} = 13.221). The effect of baseline response time was also significant (F[1,171] = 151.376, p < .001) with strong evidence in support of this finding ($BF_{10} = 4.282*10^{22}$). The main effect of inhibitory control on students' problem-solving response time was not significant (F[1, 171] = 0.113, p = 0.738), with little evidence supporting this result (BF_{10} = 0.825). Further, there was an interaction between problem type and baseline response time (F[1, 171] = 6.324, p = 0.013, η^2 = 0.003).

Exploratory Analysis: Effects of Perceptual Cues by Inhibitory Control

From the pre-registered analyses, we see that this sample is high-performing without much variance in students' inhibitory control. Further, since inhibitory control was treated as a continuous variable, it was included in our pre-registered analyses as a covariate rather than a between-subjects factor. Therefore, to explore whether students with different levels of inhibitory control in our high-performing sample may be impacted differently by perceptual cues, we conducted a median split (Mdn = .17) to compare students with higher inhibitory control (n=87) to students with lower inhibitory control (n=87). This approach allowed us to explore how students with different degrees of inhibitory control are impacted by perceptual cues in order-of-operations problems on both accuracy and response time within our high-performing sample.

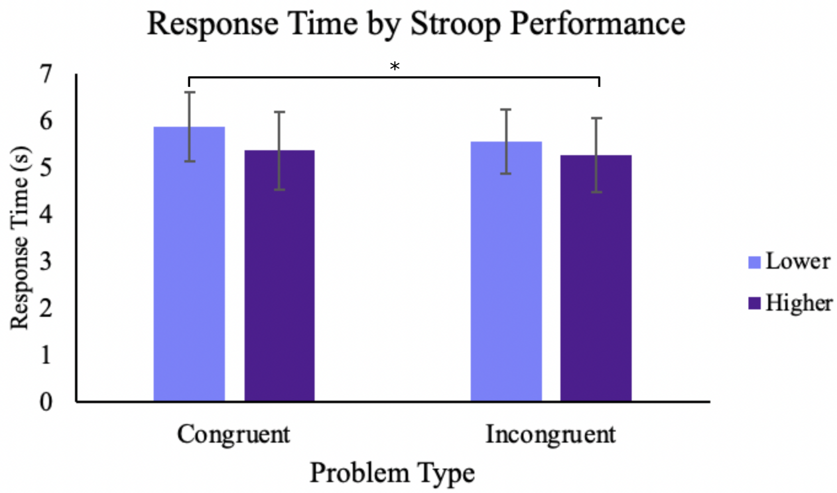
[1] By treating inhibitory control as a continuous covariate within ANOVA models, we were unable to test for interaction effects in these models using a Bayesian ANOVA that would be comparable to those presented.

A 2 (Problem Type) × 2 (Inhibitory Control: Higher or Lower) repeated measures ANOVA comparing occurring students' accuracy on order-of-operations problems revealed no significant effect of problem type ($F[1,172] = 0.068$, $p = .795$), inhibitory control ($F[1,172] = 0.021$, $p = .886$) or Problem Type × Inhibitory Control interaction ($F[1,172] = 1.428$, $p = .234$), suggesting that students were comparably accurate on all problems. A Bayesian repeated measures ANOVA revealed moderate evidence in support of the null effects of problem type ($BF_{10} = 0.122$) and inhibitory control ($BF_{10} = 0.170$).

Next, a 2 (Problem Type) × 2 (Inhibitory Control) repeated measures ANOVA comparing students' response times on order-of-operations problems revealed a significant effect of problem type ($F[1,172] = 10.146$, $p = .002$, $\eta^2 = 0.005$), confirming that students had slower response times on congruent than incongruent problems. It also revealed a marginally significant effect of Inhibitory Control ($F[1,172] = 3.365$, $p = .068$, $\eta^2 = 0.017$) with no Problem Type × Inhibitory Control interaction ($F[1,172] = 2.303$, $p = .131$), suggesting that students with lower inhibitory control had slower response times than students with higher inhibitory control (Figure 5). A Bayesian repeated measures ANOVA revealed strong evidence in support of the effects of problem type ($BF_{10} = 13.377$) and anecdotal evidence in support of the effect of inhibitory control ($BF_{10} = 1.135$). Post hoc comparisons with Bonferroni corrections revealed that students with lower inhibitory control had significantly slower response times ($M = 5.88$, $SD = 1.48$) on congruent problems than incongruent problems ($M = 5.55$, $SD = 1.38$; $p = .006$). Further students with lower inhibitory control had significantly slower response times on congruent problems than students with higher inhibitory control on incongruent problems ($M = 5.26$, $SD = 1.57$; $p = .044$).

Figure 5

Response Times on Congruent and Incongruent Problems by Inhibitory Control



Note: Error bars represent the standard deviation.
* $p < .05$ ** $p < .01$

Discussion

In the current study, we tested the impact of perceptual cues and inhibitory control on college students' performance solving order-of-operations problems. The results indicated three main findings. First, students demonstrated comparable accuracy on congruent and incongruent order-of-operations problems. Second, students demonstrated longer response times on congruent problems than incongruent problems. Exploratory analyses suggest that this effect may be driven by students with lower baseline performance who had the slowest response times on congruent problems. Third, on average, students' inhibitory control did not impact their performance on order-of-operations problems. However, exploratory results revealed that students with lower inhibitory control demonstrated significantly longer response times on congruent problems than higher-inhibitory control students on incongruent problems. In the following sections, we discuss our interpretations of each result to provide implications for researchers and educators.

Effect of Physical Spacing on Performance

Through this study, we aimed to conceptually replicate prior findings on the effects of physical spacing on college students' performance simplifying arithmetic expressions. Prior work has demonstrated that college students show higher accuracy simplifying arithmetic expressions with congruent and neutral spacing than expressions with incongruent spacing and comparable response times on both problem types (e.g., Landy & Goldstone, 2010). Here, we found two different results: students were comparably accurate across problem types but they were slower to answer congruent problems as opposed to incongruent problems. We reason that these findings may be due to one or more of the following factors: a) the high-performing sample, b) the study design, and c) a hierarchy of attentional cues.

First, the study sample may be driving the null effect of problem type on students' accuracy. Although research has shown that both children and adults are susceptible to perceptual cues and rely on perceptual-motor skills (e.g., Braithwaite et al., 2016; Gomez et al., 2014; Landy & Goldstone, 2010), it is possible that the content used in the study stimuli (i.e., arithmetic expressions with three numbers and two operands) were too simple for our sample of college students. Students were very high-performing across the board so perhaps there was not enough variance in the sample to see any effects of physical spacing in problems on students' accuracy. Notably, we did check the accuracy rates on each problem and found that students were comparably accurate on problems across problem type, suggesting that the difficulty of problems was evenly spread across problem types and that this result is not a reflection of the study stimuli.

Second, a similar explanation may also explain why students had slower response times on congruent than incongruent problems. Since research has shown that reliance on perceptual cues increases with experience and expertise (Braithwaite et al., 2016; Rumelhart, 1986), we posit that students in our high-performing sample were likely to rely on perceptual cues; further, the study design (i.e., interleaving congruent and incongruent problems) may have primed students inadvertently to exercise inhibitory control while problem solving. If students were attending to perceptual cues throughout the experimental problem block, and realized that some problems contained incongruent spacing, they may have engaged in a one- or two-step process to inhibit their impulse to calculate before correctly simplifying the expression. For example, if students were attending to the perceptual cues in an incongruent problem, they would just need to suppress their initial instinct to then solve the problem. If they next viewed a congruent problem next, they may have been primed to suppress their initial impulse to calculate followed by a realization that their initial impulse was correct, potentially taking longer to solve the problem. This rationale is well-aligned with the negative priming paradigm which posits that if an individual views a stimulus that is to be ignored, followed by a stimulus that is not to be ignored, their accuracy and response time may suffer on the latter task (Neil, 1977; Frings et al., 2015). Importantly, many prior studies that have found an effect of physical spacing on students' performance in arithmetic has applied a between-subjects design or blocked experimental problems by spacing type (e.g., Braithwaite et al., 2016; Harrison et al., 2020). Interestingly, in their "Experiment 1", Landy and Goldstone (2010) also interleaved problems at random and did not find an effect of spacing on college students' accuracy so this finding conceptually replicates their research and suggests that interleaving problems with different perceptual cues may impacts students' performance differently than blocking problems.

Third, we posit that the effect of physical spacing on students' response times may be potentially explained by an implicit hierarchy of how we attend to perceptual cues. Specifically, we consider that students may have demonstrated quicker response times on incongruent problems because the physical spacing in incongruent problems is more salient than the physical spacing in congruent problems, making them easier to perceive and process. Individuals naturally attend to multiplication operands quicker than addition operands (Landy et al., 2008) so, as discussed in Harrison et al. (2020), congruent spacing between numbers and operands may provide perceptual cues that are redundant to the notation already provided, rendering them less noticeable. If that is the case, students may have been quicker to process incongruent problems than congruent problems, leading to the difference in response times.

Inhibitory Control, Physical Spacing, and Mathematics Performance

In the current study, we found minimal evidence to suggest a relation between students' inhibitory control and performance on order-of-operations problems. In the pre-registered analyses, we found no effect of inhibitory control on students' accuracy or response time when also controlling for their baseline performance. In the exploratory analyses, we found that students with lower inhibitory control demonstrated significantly longer response times on congruent problems than higher-inhibitory control students on incongruent problems. Aligned with the explanations presented above, we interpret this finding to mean that students with lower inhibitory control (i.e., lower relative to this high-inhibitory control sample) may struggle with the two-step process to a) initially suppress the urge to calculate congruent problems and then b) process the problem to correctly perform calculations, resulting in longer response times. These findings contribute to the body of mixed results on the relation between inhibitory control and mathematics performance (Bull & Lee, 2014) by providing findings from a college-aged sample using the Animal Stroop task and novelly investigating the relation between inhibitory control and physical spacing in mathematics problems.

Limitations and Future Directions

We acknowledge that this study contained multiple limitations. First, our sample displayed high performance on both the inhibitory control task and the order-of-operations problems. The math content used in the study stimuli would be more appropriate for younger students in pre-algebra or early algebra courses. Similarly, we would expect to see more variance in performance on the Animal Stroop task with a younger sample as well since inhibitory control develops with cognitive development.

Second, the study design may have impacted results. In particular, students' response time was measured from the time that each problem was displayed on-screen until students entered their answers to the order-of-operations problems on their keyboard *and* clicked the "Next" button on-screen to advance. This two-step process may have added variance to response times rather than selecting a multiple-choice answer or pressing "return" after typing responses. As discussed above, interleaving the experimental problems rather than blocking them may have also influenced students' performance on the task by switching between viewing congruent and incongruent problems at random. Additionally, presenting the Stroop task prior to the arithmetic problems may have inadvertently primed students to suppress initial impulses while problem-solving. Future research may consider presenting experimental problems in blocks and presenting the inhibitory control task at the end of the study to alleviate the risk of priming students.

Looking ahead, we plan to conduct this study with grade school students in pre-algebra and Algebra I courses to see whether these effects replicate across sample populations or if younger students may be impacted differently by both perceptual cues and inhibitory control. Additionally, future research may consider using other measures of inhibitory control to see whether these findings may be task-specific. Further, future research may consider conceptually replicating this work with other perceptual cues (e.g., color, spatial arrangement) to see how the effects of perceptual cues generalizes.

Conclusions

This study tested the effect of physical spacing as a perceptual grouping mechanism in order-of-operations problems and whether students' inhibitory control impacted performance. We found that all college students were high-performing on both the inhibitory control measure as well as the order-of-operations problems. Students were comparably accurate on congruent and incongruent problems although they were quicker to answer incongruent problems. There was no main effect of inhibitory control on performance. By testing the association between students' performance on the animal-size Stroop task and the order of operations problems, these findings advance our understanding of the relation between inhibitory control and mathematical performance. Together, these results bridge theory on perceptual learning and cognitive developmental work by exploring how different students are impacted by perceptual cues in online mathematics activities.

References

Adkins, M., & Noyes, A. (2018). Do advanced mathematics skills predict success in biology and chemistry degrees?. *International Journal of Science and Mathematics Education*, *16*(3), 487-502.

Alibali, M. W., Crooks, N. M., & McNeil, N. M. (2018). Perceptual support promotes strategy generation: Evidence from equation solving. *British Journal of Developmental Psychology*, *36*, 153–168. https://doi.org/10.1111/bjdp.12203

Bellon, E., Fias, W., & De Smedt, B. (2016). Are individual differences in arithmetic fact retrieval in children related to inhibition?. *Frontiers in Psychology*, *7*, 825.

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.

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, *8*(1), 36-41.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*(3), 273-293.

Cassotti, M., Agogué, M., Camarda, A., Houdé, O., & Borst, G. (2016). Inhibitory control as a core process of creative problem solving and idea generation from childhood to adulthood. *New Directions for Child and Adolescent Development, 151,* 61-72.

Chan, J. Y. C., Sidney, P. G., & Alibali, M. W., (2019, November). Corresponding color coding facilitates learning of area measurement. In Otten, S., Candela, A. G., de Araujo, Z., Haines, C., & Munter, C. (Eds), *Proceedings of the forty-first annual meeting of the*

North American Chapter of the International Group for the Psychology of Mathematics Education. St Louis, MO: University of Missouri. (pp. 222-223).

Closser, A. H., Chan, J. Y. C., Smith, H., & Ottmar, E. (under review). Perceptual learning in math: Implications for educational research, practice, and technology. Rapid Community Report Series. Digital Promise and the International Society of the Learning Sciences.

Common Core State Standards (2010). Common Core State Standards for Mathematics (CCSSM). *Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers*.

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.

Faulkenberry, T. J., Ly, A., & Wagenmakers, E. J. (2020). Bayesian inference in numerical cognition: A tutorial using JASP. *Journal of Numerical Cognition*, *6*(2), 231-259.

Frings, C., Schneider, K. K., & Fox, E. (2015). The negative priming paradigm: An update and implications for selective attention. *Psychonomic bulletin & review*, *22*(6), 1577-1597.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. East Norwalk, CT: Appleton-Century-Crofts.

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *ZDM*, *47*(5), 771-782.

Goldstone, R. L. (1998). Perceptual learning. *Annu. Rev. Psychol*, *49*, 585-612.

Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science, 2*(2), 265-284. https://doi.org/10.1111/j.1756-8765.2009.01055.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.

Gómez, D. M., Benavides-Varela, S., Picciano, C., Semenza, C., & Dartnell, P. (2014). Symbol spacing and its influence in school children's calculation of mixed-operation arithmetic expressions. In Oesterle, S., Nicol, C., Liljedahl, P., & Allan, D. (Eds.). *Proceedings of the Joint Meeting of PME 38 and PME-NA 36*, Vol. 6, pp. 310. Vancouver, Canada: PME.

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.

Hoch, M. & Dreyfus, T. (2004). Structure sense in high school algebra: The effect of brackets. In M. J. Hoines & A. B. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, Vol 3 (pp 49- 56). Bergen, Norway: Bergen University College.

JASP Team (2020). JASP (Version 0.14.1)[Computer software].

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.

Jupri, A., Sispiyati, R., & Chin, K. E. (2021). An investigation of students' algebraic proficiency from a structure sense perspective. *Journal on Mathematics Education*, *12*(1), 147-158.

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

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., ... & Velez, E. D. V. (2015). *The Condition of Education 2015*. NCES 2015-144. *National Center for Education Statistics*.

Kirshner, D. (1989). The visual syntax of algebra. *Journal for Research in Mathematics Education, 20,* 274–287. https://doi.org/10.2307/749516

Kroesbergen, E. H., Van Luit, J. E. H., Van Lieshout, E. C. D. M., Van Loosbroek, E., & Van de Rijt, B. A. M. (2009). Individual differences in early numeracy: The role of executive functions and subitizing. *Journal of Psychoeducational Assessment, 27*(3), 226-236.

Landy, D., & Goldstone, R. L. (2007). How abstract is symbolic thought?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 720.

Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *Quarterly Journal of Experimental Psychology*, *63*(10), 1953-1968.

Lawson, A., Ayala, B., & Son, J. Y. (2021). Priming students to calculate inhibits sense-making. *Journal of Cognitive Science*, *22*(1), 41-69.

Lee, J. E., Hornburg, C. B., Chan, J. Y. C., & Ottmar., E. (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 patched? *Child Development Perspectives, 13*(1), 28–33. https://doi.org/10.1111/cdep.12304

Livneh, D., & Linchevski, L. (2007, July). Algebrification of arithmetic: Developing algebraic structure sense in the context of arithmetic. In *Proceedings of the 31st Conference of the Psychology of Mathematics Education* (Vol. 3, pp. 217-225).

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), 25.

Neill, W. T. (1977). Inhibition and facilitation processes in selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 444–450.

Papadopoulos, I., & Gunnarsson, R. (2020). Exploring the way rational expressions trigger the use of "mental" brackets by primary school students. *Educational Studies in Mathematics, 103*(2), 191-207.

Parsons, S., & Bynner, J. (2005). Does numeracy matter more? London: National Research and Development Centre for Adult Literacy and Numeracy.

Rivera, J., & Garrigan, P. (2016). Persistent perceptual grouping effects in the evaluation of simple arithmetic expressions. *Memory & Cognition*, *44*(5), 750-761.

Rumelhart, D., Smolensky, P., McClelland, D., & Hinton, G. (1986). Schemata and sequential thought processes in pdp models. In *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 7–57). Cambridge: MIT Press.

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128-142.

Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex, 49*(10), 2674-2688.

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *The European Health Psychologist*, *16*(2), 75-84.

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*, 1172–1217.

Wilkey, E. D., Pollack, C., & Price, G. R. (2019). Dyscalculia and typical math achievement are associated with individual differences in number-specific executive function. *Child Development, 91*(2), 596–619. https://doi.org/10.1111/cdev.13194

Chapter 4. Viewing vs. Mirroring: The Effects of Action and Self-Explanation in Worked Examples on Algebra Learning

Whereas the previous two studies demonstrate how students are impacted by perceptual cues *during* problem-solving activities in math, the current chapter explores how perceptual and embodied supports may be integrated earlier in instructional activities to support students as they learn new concepts and procedural rules. To that end, this study investigates how students learn from studying different types of worked examples designed with an online dynamic notation tool.

This study was planned and conducted in collaboration with Hannah Smith, Jenny Yun-Chen Chan, and Erin Ottmar over the past couple of years. I developed and refined this project plan through multiple iterations of presentations and small student research grant applications between 2020-2021; ultimately, I was successfully awarded a Psi Chi Graduate Research Grant to provide participant compensation. In light of the challenges faced during the COVID-19 pandemic, recruitment proved to be incredibly difficult. With that in mind, this chapter presents preliminary results from this study and discusses future directions and implications of this research.

Abstract

Worked examples are effective instructional support for algebra learning, including when they are paired with self-explanation prompts (Booth, et al., 2013; Renkl, 2014). Their effectiveness is predominantly explained by cognitive load theory; specifically, worked examples offload strains on students' cognitive capacities and free up working memory to support learning. Alternatively, I propose that incorporating principles of embodied cognition into worked examples may positively impact learning even if the worked examples do not reduce students' cognitive load. Building on these theories, this study investigates how to most effectively present worked examples in online learning environments. I leveraged the affordances of a dynamic algebra notation tool to test how student actions and self-explanations impact learning in different worked example formats. I predicted that algebra students may learn more when they dynamically *mirror* worked examples on-screen, rather than simply *view* worked examples. A total of 64 ninth-grade Algebra I students completed a three-day online study that included an intervention in which students were randomly assigned to: a) *view,* b) *view-and-explain*, c) *mirror*, or d) *mirror-and-explain* worked examples and complete paired practice problems. Results from this sample suggest that, on average, all students experienced learning gains after participating in the intervention. Further, students who received self-explanation prompts with worked examples experienced larger gains than their peers. These findings provide preliminary evidence in support of prior research on the worked example effect and suggest that researchers, teachers, and content developers may consider using alternative worked example formats in online learning environments for algebra.

Keywords: worked examples, algebra, educational technologies, cognitive load theory, embodied cognition

Viewing vs. Mirroring: The Effects of Action and Self-Explanation in Worked Examples on Algebra Learning

Mathematics skills and achievement are related to significant outcomes later in life such as academic achievement and employment (Adkins & Noyes, 2018; Lee, 2013; Parsons & Bynner, 2005; VanDerHeyden & Burns, 2009); however, many students struggle to progress beyond Algebra I (Kena et al., 2015). To support student reasoning and learning in algebra, cognitive theories can be leveraged to design effective, evidence-based instructional support for students. Worked examples, which provide students with a step-by-step solution to a given problem, are an effective instructional support that have been widely used across subjects (e.g., chemistry: McLaren et al., 2016; computer science: Zhi et al., 2019; physics: Chi et al., 1989), including algebra (Booth et al., 2013; Booth et al., 2015; Carroll, 1994; Foster et al., 2018). Studying worked examples in algebra leads to more efficient student learning and higher learning rates than solely working through problems without guided support (e.g., Barbieri & Booth, 2020; Booth et al., 2013; Carroll, 1994). This phenomenon is known as the worked example effect (Sweller, 2006).

Although extensive research has shown that studying worked examples positively impacts learning in algebra (Barbieri & Booth, 2020; Booth et al., 2013; Booth et al., 2015 Carroll, 1994; Foster et al., 2018), the *features* of worked examples that make them effective are still largely unknown. While some researchers have examined worked examples through cognitive load theory (e.g., Chandler & Sweller, 1991; Sweller, 1988, 1989), researchers have examined worked example through other theoretical perspectives as well (Renkl, 2014), such as analogical reasoning (e.g., Nokes-Malach et al., 2013), social cognition (e.g., van Gog & Rummel, 2010), and observational learning (Bandura, 1986; Zimmerman & Kitsantas, 2002). Here, I contribute to the larger literature on example-based learning by examining the impacts of worked examples in learning of algebraic equation solving through the cognitive load theory as well as perceptual learning theory and self-regulated learning. This work may advance our understanding of when and why worked examples support learning, and how the design of worked examples could be informed and improved upon by using multiple theoretical perspectives.

Additionally, studying worked examples online decreases instructional time and increases student learning (Salden et al., 2010), and some studies have examined the effectiveness of integrating worked examples in Cognitive Tutor (Reed et al., 2013) or embedding videos in an online learning environment (Hoogerheide et al., 2019; van Gog et al., 2011). Although worked examples of mathematical problems have been studied in online environments, they are still typically displayed as static texts and images or as videos of a person modeling and rewriting derivations. With the advancement of educational technologies, online worked examples can be presented in more dynamic ways demonstrating the problem-solving processes through fluid transformations of equations. Further, dynamic technologies can also provide opportunities for students to interact with worked examples and experience the problem-solving process.

As K-12 education continues to shift towards using online resources, it is essential to consider how we can leverage different cognitive theories to design worked examples that support students as they complete assignments online. Prior research has shown the effectiveness of providing worked examples with self-explanation prompts on improving student learning in algebra (Renkl, 2002, 2014) and demonstrated that different presentations of worked examples impact student learning outcomes (Schalk et al., 2020). However, limited work has applied theories of embodied cognition to the design of worked examples in algebra or considered how self-explanations may impact learning when used with worked examples that involve student action in an interactive online setting. Here, I conduct and analyze a randomized controlled trial to test how principles of cognitive load theory and embodied cognition extend to worked examples with and without self-explanation prompts in order to identify effective ways to present algebraic worked examples in dynamic online learning environments.

Theoretical Background

Cognitive Load Theory

Worked examples have been extensively studied through the lens of cognitive load theory and the relationships between the types of cognitive load: intrinsic load (i.e., cognitive demands based on the complexity of the information being processed and the learner's prior knowledge), extraneous load (i.e., unnecessary working memory resources required for learning based on instructional approaches), and germane load (i.e., working memory resources essential to learning; Chandler & Sweller, 1991; Sweller, 1988, 1989, 1994; Sweller et al., 2019). From this perspective, when students study worked examples that lower intrinsic and extraneous load, they have more cognitive resources to draw upon in order to engage in sense-making activities that support germane load, like self-explanations, which lead to long-term learning (Leppink et al., 2013). Research on the worked example effect has shown that studying worked examples reduces students' extraneous cognitive load, leading to improvement in algebra learning more than solely practicing problems (Booth et al., 2013; Carroll, 1994; Foster et al., 2018). Based on this theory, worked examples are considered to be an effective tool for learning because students can view a step-by-step example rather than holding the pieces of information in their working memory. By offloading some of the cognitive demands of problem solving onto worked examples, students have more cognitive resources to draw connections, notice the procedural rules being applied, and construct schemas to support learning and transfer (Renkl, 2014).

That said, the presentation of worked examples has been shown to impact the effect of worked examples on learning. For example, the presentation of worked examples may impact whether students experience higher gains in procedural or conceptual knowledge (Schalk et al., 2020). Additionally, different visual features may either increase or decrease learners' extraneous and germane cognitive loads that in turn, impact learning. Specifically, Chandler and Sweller (1991) demonstrated that worked examples with information in multiple places splits learners' attention, demanding more cognitive capacity and consequently being less effective for learning than worked examples which integrate sources of information such as text and pictorial instruction. Further, Sweller and colleagues (2019) suggested that it is better to watch animations rather than static presentations to teach cognitive tasks which involve human movement. Asking learners to observe movement does not place additional extraneous cognitive load on the learner. Therefore, cognitive load theory posits that students *viewing* worked examples that minimize extraneous cognitive load by integrating instruction and movement in one place and excluding unnecessary information is likely to be most beneficial for learning.

Embodied Cognition and Design

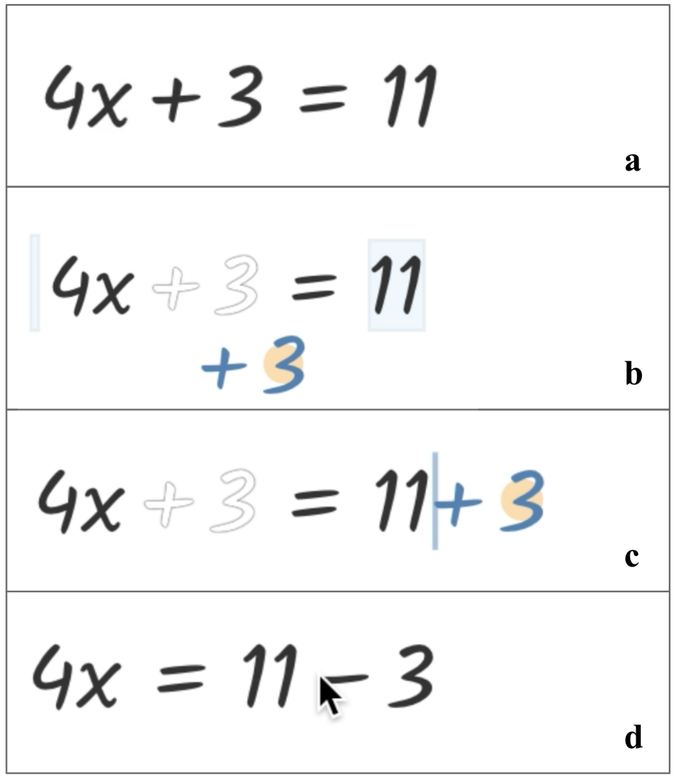
From another perspective, embodied cognition theorists may argue that students *interacting* with worked examples in an activity that leverages perception and action may be better for learning; although, limited research has considered how worked examples may be designed based on theories of embodied cognition. Theories of embodied cognition posit that thinking does not occur entirely internally, independent of the external environment; rather, students' physical experiences and interactions with their environments influence their cognitive processes, including mathematical thinking, reasoning, and learning (Barsalou, 2008; Foglia & Wilson, 2013; Lave, 1988; Nathan, 2014, 2021; Wilson, 2002). Substantial research has shown the benefits of students grounding their knowledge of mathematics concepts in embodied learning activities (see Abrahamson et al., 2020 for a review). For instance, students learn more and retain more knowledge from a mathematics lesson when they are instructed or encouraged to use purposeful gestures during the lesson (Broaders et al., 2007; Cook et al., 2008). Additionally, research has suggested that even having students observe movements can be beneficial for learning, known as the human movement effect (Sweller et al., 2019).

Substantial research has shown evidence of embodied cognition, and the benefits of students grounding their knowledge of mathematics concepts in physical experiences (e.g., Abrahamson et al., 2020). However, limited work has applied theories of embodied cognition to the design and implementation of instructional support in algebra, including worked examples of equation solving, despite a call from cognitive load theorists to integrate evolutionarily based skills (e.g., facial recognition, speaking, gesture) in instruction to lessen cognitive demands (Paas & Sweller, 2012). Notably, Ginns and colleagues (2016, 2020) demonstrated this potential synergy between cognitive load theory and embodied cognition by utilizing gestures in worked examples on angle theorems. They found that students who traced worked examples on-screen scored higher on a transfer test than their peers who solely viewed worked examples, replicating previous work in favor of tracing worked examples (Hu et al., 2015). Further, Yeo and Tzeng (2020) replicated these findings when students traced worked examples about angle relationships with parallel lines but not when students traced worked examples about laws of exponents, suggesting that the effectiveness of using embodied techniques in worked examples may be dependent on the visuospatial nature of the mathematical content. Since individuals tend to treat mathematical symbols as objects (e.g., Dörfler, 2003; Landy & Goldstone, 2010), I predict that principles of embodied cognition may apply to instructional support in algebra. By using dynamic technologies, students can interact with mathematical symbols as objects through mathematically grounded gesture-actions. Specifically, worked examples may be effective for learning when students drag and combine symbols to manipulate expressions on-screen to reproduce worked examples rather than simply viewing traditional worked examples on-screen.

Dynamic learning technologies allow users to manipulate linear equations, graphs, and expressions and see the outcomes of their actions in real-time on their computer screens. These technologies also enable teachers to present worked examples that move beyond traditional, static images by providing dynamic visual features that would not be possible to present in textbooks. The current study utilizes Graspable Math (GM; Weitnauer et al., 2016), an interactive algebra notation tool, to explore the benefits of worked examples presented with the opportunity to use dynamic gesture-actions. GM was developed from theories of perceptual learning and embodied cognition to allow users to physically manipulate mathematical notation through gesture-actions that emulate mathematical properties in a physical-to-virtual embodied experience with mathematical terms. As an example, to distribute 3 into $(2+x)$, users can touch and drag the 3 into the parentheses, automatically triggering a visualization where 3 is distributed to 2 and x, transforming the expression into $6+3x$. Through dynamic gestures, users can also combine terms, apply operations to both sides of an equation, and rearrange terms through commutative, associative, and distributive properties (Figure 1). For a review of GM and the ways that its interactivity supports intuitive, embodied, and trained perceptual-action processes, see Goldstone and colleagues (2017) and Abrahamson and colleagues (2020).

Figure 1

Example of a Dynamic Transformation Using Graspable Math



Note. Users can dynamically transform equations (e.g., 4*x*+3=11; 1a) by dragging and dropping terms (e.g., “3”; 1b, 1c) to create equivalent forms (e.g., 4*x*=11-3; 1d).

Prior work has demonstrated the efficacy of GM and its positive effect on student learning (Chan et al., 2021; Hulse et al., 2019; Ottmar et al., 2015). However, no prior research has considered using the tool as a means for students to interact with worked examples during algebra practice and potentially benefit from the automatic calculations, fluid visualizations, and feedback provided by the system. Currently, I extend the human movement effect (Sweller et al., 2019) to explore whether prompting interactions with worked examples in GM through grounded gesture-actions will increase students’ learning beyond just watching the dynamic problem-solving process of worked examples in the GM system. I suggest that these movements will not increase students’ extraneous cognitive load but may increase germane cognitive load and provide additional cues to help them process the content in the worked examples as well as generalize content by focusing on the step-by-step subgoals of worked examples during the mirroring process, in line with prior work on sub-goal learning (Margulieux et al., 2016). I posit that students may ground their knowledge of simplifying equations in the act of manipulating symbols on-screen as part of studying worked examples and may develop perceptual and procedural fluency by participating in this additional guided practice.

Worked Examples with Self-Explanations

Since worked examples provide effective instructional support for students, research over the last several decades has investigated how to best design and implement worked examples in instruction to improve learning (for a recent commentary, see Mayer, 2020). Stemming from the self-regulated learning perspective, two recommendations have emerged: incorporating paired practice problems with a worked example (Foster et al., 2018; Sweller & Cooper, 1985; Sweller, 2006; van Gog et al., 2020) and prompting students to explain the steps completed in a worked example (e.g., Aleven & Koedinger, 2002; Renkl, 2002; 2014).

First, it has been shown that worked examples are effective instructional support when students have an opportunity to practice applying their knowledge. As practice, students typically complete problems that are similar in structure to the worked examples, and they reach the solution on their own, without scaffolded support. Students studying worked examples followed by problem solving has been referred to as a Worked Examples then Problem Solving schedule. This schedule is considered to be effective because studying a worked example lowers students' extraneous load while increasing germane load to allow for schema acquisition and knowledge that can then be applied to the following practice problem. Worked examples followed by problem solving has been shown to improve student learning and transfer beyond problem solving alone (Carroll, 1994; Cooper & Sweller, 1987; Retnowati et al., 2010; Rourke & Sweller, 2009; van Harsel, et al., 2020; Ward & Sweller, 1990). Recent research has also revealed that providing practice problems followed by worked examples may be just as effective for learning as studying worked examples followed by problem solving (van Harsel et al., 2019, 2020). Further, when students have the autonomy to regulate their own learning, they tend to alternate between practice problems and worked examples so they can understand the limit of their knowledge, then effectively learn from the examples (Foster et al., 2018). In sum, prior research has led to extensive evidence in favor of using paired worked examples and practice problems in instructional activities and for that reason, I include paired practice problems with the worked examples in our current research.

Second, studying worked examples with self-explanation prompts increases learning above and beyond studying worked examples alone (e.g., Aleven & Koedinger, 2002; Berthold et al., 2009; Nokes-Malach et al., 2013; Renkl, 2002; 2014). As students explain worked examples, they may engage in deductive processes, generalization, and making the implicit knowledge explicit (Chi & VanLehn,1991), and the process of self-explanation helps students monitor their understanding and regulate their learning (VanLehn et al., 1992; see Renkl & Eitel, 2019, for a review). Research has shown that eliciting self-explanations is moderately effective for learning, particularly in domains with general principles that guide problem solving (Rittle-Johnson & Loehr, 2017), and even recommended self-explanation for instructional practice (e.g., Chi et al., 1989; Chi et al., 1994; see Dunlosky et al., 2013 for a brief review). Self-explanation prompts in worked examples leverage this effect by challenging students to reflect on, make meaning of, and articulate the content they study in a worked example. Prompting self-explanations typically increases students' cognitive load by presenting an additional challenge to reflect on instructional content (e.g., Hilbert & Renkl, 2009) so in order for self-explanations to be effective for learning when paired with worked examples, they should be implemented to minimize extraneous load (Renkl & Eitel, 2019). In particular, principle-based self-explanations prompt students to connect conceptual knowledge to procedural rules and practices in order to help with future problem-solving, potentially through analogical reasoning and noticing similarities between problem structures (Renkl, 2014; Renkl & Eitel, 2019).

However, limited work has examined how the presentation of worked examples with self-explanations may impact learning in interactive online settings. On the one hand, self-explanations may place additional cognitive demands on students which could detract from the effectiveness of a worked example. Similarly, Renkl (2014) posited that self-explanation prompts are not beneficial beyond studying worked examples when students are faced with complex learning tasks that induce a lot of cognitive load, such as problems with high element interactivity (the number of elements that must be held in working memory for a given problem; Chandler & Sweller, 1996). On the other hand, self-explanations have been shown to add value to worked examples for both correct and incorrect worked examples, suggesting that the self-explanation effect is somewhat robust (e.g., Aleven & Koedinger, 2002; Barbieri & Booth, 2016, 2020; Hilbert et al., 2008). More research is needed to determine whether and how the presentation of worked examples impacts the effect of self-explanations on learning.

I posit that the impact of self-explanation prompts in worked examples may be contingent on the presentation of the worked example itself and the theories which inform the design of the worked example. Specifically, worked examples designed from cognitive load theory intentionally minimize extraneous information and integrate visual features to reduce students' cognitive load while viewing the worked example, freeing up cognitive capacity to make connections and provide self-explanations that lead to learning beyond simply studying worked examples. From this viewpoint, if worked examples are designed from another theory such as embodied cognition, that require actively mirroring worked examples and consequently may involve a higher level of element interactivity per example, students may experience cognitive overload when asked to also provide self-explanations (Renkl, 2014). Conversely, such worked examples may benefit from having self-explanation prompts so that students reflect on the purpose of, and mathematical principles behind, each step rather than simply following the steps in the worked examples instructions. Since self-explanation prompts have been effective for learning from viewing worked examples (Aleven & Koedinger, 2002; Barbieri & Booth, 2016, 2020; Hilbert et al., 2008), I posit that this effect may extend to worked examples that involve student action and seek to replicate previous research on the added impact of self-explanations in worked examples.

Testing and Integrating Cognitive Theories to Design Worked Examples

While cognitive load theory supports the use of worked examples that minimize extraneous load, perception and action may also play an important role in learning through worked examples. Notably, prior work has shown that incorporating student actions, such as tracing, while studying worked examples on geometry leads to decreased extraneous load and higher performance on recall and transfer tests than peers who simply viewed worked examples (Ginns et al., 2016, 2020; Tang et al., 2019; Yeo & Tzeng, 2020). Currently, I seek to extend this line of inquiry to explore how interactive worked examples presented in a dynamic online environment influence student learning and cognitive load in the context of algebra.

Based on the bodies of literature surrounding worked examples, embodied cognition, and self-explanations in instructional materials, I predict that worked examples may be more effective for learning when students are active participants and ground their knowledge in embodied experiences, such as using movement for problem-solving. Specifically, the benefit of mirroring steps of a worked example through dynamically transforming expressions on-screen may outweigh the risks of presenting potentially extraneous information and demanding additional actions. For example, Reed and colleagues (2013) found that students who studied worked examples with interactive graphics outperformed their peers who studied static-table and static-graphic worked examples. Based on their findings, Reed and colleagues (2013) suggested that students who studied interactive worked examples might have benefitted more if those examples also included self-explanation prompts to challenge students to make connections between their actions and the mathematical concepts. Therefore, I predict that interactive worked examples, such as mirroring steps to solve an equation on-screen, would be more effective for learning when paired with self-explanations. Additionally, while interactive worked examples may be equally or more effective for learning than simply viewing worked examples, I anticipate that measuring students' levels of cognitive load will provide further insights into the relation between the different presentations of worked examples and student learning.

The Current Study

Since extensive evidence demonstrates the efficacy of studying worked examples over practicing problem-solving alone for skill acquisition (e.g., Barbieri & Booth, 2020; Booth et al., 2013; Carroll, 1994; Renkl, 2014, 2017), the current study aims to extend this research by investigating how student action and self-explanations impact learning through worked examples in a dynamic online learning environment. I designed and deployed a randomized controlled trial for Algebra I students. Using a 2 (Presentation: viewing vs. mirroring) × 2 (Self-Explanation: self-explanation prompts vs. no self-explanation prompts) between-subjects design, I examined the effects of mirroring the steps within worked examples, explaining those steps, and their interaction on student learning. Specifically, students completed an online algebra activity in which they were assigned to either: a) *view,* b) *view-and-explain*, c) *mirror*, or d) *mirror-and-explain* worked examples and complete paired practice problems. Immediately after completing the intervention, students also completed a measure of cognitive load (adapted from Leppink et al., 2013) that identifies self-reported levels of intrinsic, extraneous and germane cognitive load. In addition to observing differences in learning by condition, this measure allows us to tease apart the effect of each worked example condition on each facet of students' cognitive load. I pose the following questions:

1. *Do students learn more from viewing or mirroring guided worked examples in an online learning environment, regardless of self-explanation prompts?*
2. *Do students learn more from studying worked examples with or without self-explanation prompts, regardless of whether they view or mirror worked examples?*
3. *Is there an interaction between worked example presentation and self-explanation prompts? (i.e., is mirroring most effective when paired with self-explanation prompts?)*
4. *Do students report different levels of cognitive load after studying different types of worked examples?*

First, I hypothesize that algebra students may learn more by *mirroring* worked examples as they manipulate terms on-screen to reproduce steps of worked examples with GM rather than simply *viewing* worked examples on-screen. Specifically, mirroring the worked examples may help students take the time and actions to make connections that are not apparent just by viewing worked examples. Instead, as students experience the problem-solving process themselves, they may gain an understanding of how their actions lead to mathematical results. Second, I hypothesize that I will replicate prior findings that students learn more from studying worked examples with versus without self-explanation prompts. Consequently, I also hypothesize that self-explanations provide an added benefit beyond mirroring worked examples such that mirroring worked examples may be most effective for learning when students are prompted for self-explanations. Last, analyzing students' self-reported levels of cognitive load may reveal how studying the four different worked example types are related to levels of intrinsic, extraneous, and germane load to tease apart the mechanisms through which worked examples influence learning. I hypothesize that students in the mirror conditions may report higher levels of germane load.

Methods

Participants

Full data collection occurred from September 2021-February 2022. I recruited seventh- to ninth-grade Algebra I students through an existing pool of teachers who already use GM as well as through social media outlets, local teaching communities, and reaching out to local school districts. To thank participating classes, teachers received a $50 gift card for classroom supplies.

Three teachers assigned the study to their students; however, students from one of the teachers were excluded due to only completing one of the three days of the study at random rather than completing each of the three days in succession. As a result, 64 students from two teachers were included in our current analytic sample. These students were included in the analytic sample because they completed at least four of the eight problems on the pretest on Day One, participated in the intervention for over 10 minutes on Day Two (i.e., suggesting that they completed the brief tutorial and studied at least one worked example), *and* finished at least four of the eight problems on the posttest on Day Three. I decided to include students who completed at least half of the pretest and posttest because the assessments contained four near-transfer and four far-transfer items; this allowed us to retain a decent sample size while being able to approximate students' performance on the assessments.

All 64 students were in ninth grade, Algebra I courses ranging from 13-15 years old (M = 14.19 years, SD = .47). Our sample included: 34 (53%) female, 27 (42%) male, and two (3%) non-binary students, with one student electing to not report gender. Further, students provided whether they had any prior experience with GM: 30 (47%) of students responded "No", 25 (39%) responded "Yes", and 9 (14%) students were unsure whether they had any prior experience using GM.

Study Procedure

This study procedure received approval by our institutional review board prior to data collection. The study was conducted online in students' web browsers through Qualtrics and Graspable Math Activities, an extension of GM that allows teachers to create and assign activities for students (https://activities.graspablemath.com). Teachers assigned the study as an in-class activity for students to complete in three 20- to 30-minute sessions in three consecutive class periods. Students were instructed to work individually on their own devices at their own pace to complete the study. Students were allowed to use writing utensils and scrap paper but no calculators as they completed the study. While teachers assigned the activity to their students and provided a link to the study, all other directions for students were within the online assignments to minimize the variability in the fidelity of implementation by teachers.

On Day One, students were directed to Qualtrics to complete an eight-item pretest on equation-solving to measure their algebra knowledge. On Day Two, students opened a Qualtrics survey that randomized each student into one of four conditions and automatically redirected the student to Graspable Math Activities within their assigned condition. In Graspable Math Activities, all students completed a brief (approximately five-minute) training tutorial. Once students completed the tutorial, they completed four pairs of worked examples and practice problems. The worked example formats and instructions varied across conditions. The worked examples and paired practice problems were presented in the same order across the four conditions, alternating between a worked example followed by a paired practice problem. Immediately after completing the four pairs of worked examples and practice problems, students completed a ten-item survey to measure their cognitive load. On Day Three, students completed an eight-item posttest followed by a brief demographics survey administered in Qualtrics. The materials and measures for each day are described below.

Day One: Pretest

Algebra Knowledge Assessments

An eight-item pretest measured students' algebra knowledge at baseline. The pretest contained six open-source problems (adapted from Engage NY and Project Utah curricula) and two items with similar equation structures that were designed by our project team (Appendix A). Four of the items match the equation structures used in the worked examples (items 1, 4, 7, 8) and the other four items are transfer problems (items 2, 3, 5, 6). For each problem, students were instructed to, "*Solve the following equation. Enter the value of x below as a whole number or fraction*."

These assessments were used in a previous study with a similar population (Smith et al., 2022); the reliability coefficient of these eight items was KR-20 = 0.86 at pretest and KR-20 = 0.89 at posttest, showing high internal consistency across items. Student performance was not at floor or ceiling for any of the items in the previous study, suggesting that the assessments would be appropriate measures of algebra knowledge for this study. With the current sample, the reliability coefficient was KR-20 = .74 at pretest and KR-20 = .77 at posttest. The lower reliability coefficients in the current study may be explained by the smaller sample size since the samples were from comparable populations; however, the reliability on posttest items still indicates a positive correlation among the test items.

Day Two: Intervention

Tutorials

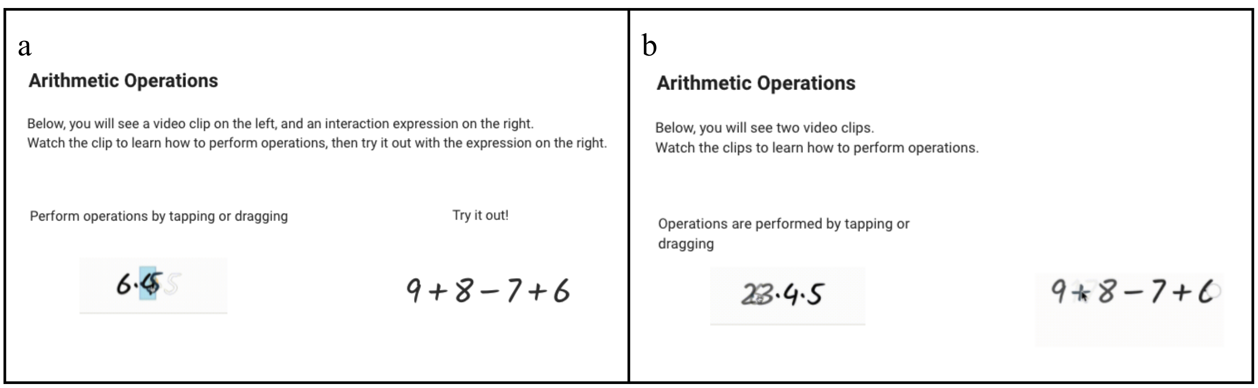
On Day Two, students were randomly assigned to one of the four conditions as they opened the assignment link. Based on their assigned condition, students completed one of two tutorial trainings prior to starting the intervention. Specifically, students in the *mirror* and *mirror-and-explain* conditions viewed six looping videos (one at a time) demonstrating the gesture-actions in GM for (1) arithmetic operations (e.g., $2 \cdot 3 \cdot 4 \cdot 5$), (2) commuting a term (e.g., moving 2 from the left to right side of the expression (e.g., transforming 2+4 into 4+2), (3) commuting a variable with a coefficient (e.g., $3x$), (4) distributing a number into parentheses (e.g., transforming $2(3+2x)$ into $6+4x$), (5) performing inverse operations such as moving a number from one side of the equals sign to the other, and 6) performing inverse operations with multiplication and division on both sides of an equation. For example, to perform multiplication in $2 \cdot 3 \cdot 4 \cdot 5$, the video demonstrated tapping the multiplication dot between 2 and 3 to transform the expression into $6 \cdot 4 \cdot 5$. The video then showed students that they can also drag 5 on top of 4 to multiply the numbers and transform the expression into $6 \cdot 20$. Next to each demonstration video, students were provided with a structurally similar expression (e.g., 9+8−7+6) to practice each action using GM, similar to what they were instructed to do in the intervention (Figure 2a).

At the end of the tutorial, students saw an interactive equation 2(x+3)=20, and used the learned gesture-actions to solve for x.

Students in the *view* and *view-and-explain* conditions completed a similar tutorial, but they did not use GM to perform actions on the practice expression. Instead, each practice expression was replaced with a demonstration video, so students in the viewing conditions saw the same transformations that students in the mirroring conditions saw but did not have the opportunity to interact with the expressions (Figure 2b). Specifically, students saw two videos demonstrating each action. At the end of the tutorial, they viewed a video of the equation-solving process for 2(x+3)=20 and were asked to enter the solution.

Figure 2

Tutorial example in the (a) mirroring conditions and (b) viewing conditions.



Worked Examples and Paired Practice Problems

After the tutorial training, all students completed four pairs of worked examples and practice problems. The worked examples and practice problems were used in a previous study on worked examples (Smith et al., 2022), and were the same across the four conditions. The four conditions varied in (1) whether students viewed or mirrored the worked examples, and (2) whether students provided self-explanation of the worked examples. For each worked example and practice problem pair, students first studied the worked example as directed by the instructions displayed for their condition. Even while keeping the domain content the same, different implementations of worked examples in instructional activities can lead to differences in time on task among students (Zhou et al., 2015). In this case, the time spent on studying worked examples may vary systematically by condition (i.e., students in mirror-and-explain condition may naturally spend more time on each worked example compared to students in view condition due to the different demands within each condition) so I used a timer in the GM system to ensure that all students in all conditions spent a minimum of two minutes on each worked example to minimize differences in time-on-example between conditions, similar to prior work on example-based learning (Ginns et al., 2016).

After studying the worked example, students solved the paired practice problem on the next screen. Each practice problem displayed a similar equation structure as the worked example prior but without a worked example in view to reference. The equations and derivations used for the worked examples were developed by Rittle-Johnson and Star (2007), and used in their previous work on worked examples (Appendix B). I created the four paired practice problems to match the structure of each worked example. For each practice problem, students solved for the variable in the equation and entered their answer as a number.

Additionally, in the explain conditions (i.e., view-and-explain, mirror-and-explain), I presented lowly-structured, principle-based self-explanation prompts (Renkl & Eitel, 2019) that encouraged students to connect the observed procedures in the worked examples with the underlying mathematical principles that are behind each derivation step. These open-ended prompts provided students opportunities to make implicit knowledge explicit while keeping the experimental session within 30 minutes, feasible for in-class data collection. Although the typical practice is to provide students with correct explanations prior to self-explanations prompts (e.g., Rittle-Johnson & Loehr, 2017; Renkl & Eitel, 2019), I decided to only provide the self-explanation prompts in order to maintain the feasibility of the study. Because students in the mirror-and-explain condition were already studying the worked examples, reproducing them using GM, and explaining their steps, I wanted to avoid overwhelming them with studying correct explanations, and potentially lengthening Session 2 beyond 30 minutes. Prompting students to explain the steps within each worked example is also aligned with prior work that demonstrates the benefits of having students self-explain subgoals within problems (Margulieux et al., 2016).

Conditions

This study used a 2 (Presentation: view or mirror) ✕ 2 (Self-explanation: prompt or no prompt) between-subjects design to examine the effect of four experimental conditions on student learning. Students were randomly assigned to one of the four conditions in which they would: 1) *view* worked examples, 2) *view-and-explain* worked examples, 3) *mirror* worked examples, or 4) *mirror-and-explain* worked examples as they completed the assignment (Figure 2).

View Condition. For each worked example, students in the *view* condition saw two presentations of a worked example on-screen. First, the worked example image on the left side of the screen displayed the major derivations of each problem in a static image, modeled from derivations used in prior research (Rittle-Johnson & Star, 2007). Second, the worked example on the right side of the screen displayed a looping video of the problem being transformed in GM. I included the looping video so that students in the view and mirror conditions both saw the dynamic problem-solving process; the only difference is that students in the mirror conditions generated the problem-solving process themselves through gesture-actions. In the view condition, students were prompted to "*Study the worked example. Once you feel comfortable with the steps taken to solve for the variable, select the solution below as your answer.*" (Figure 3a). I included the multiple choice question asking students to select the solution of the equation in order to ensure that the students in the view condition studied the worked examples.

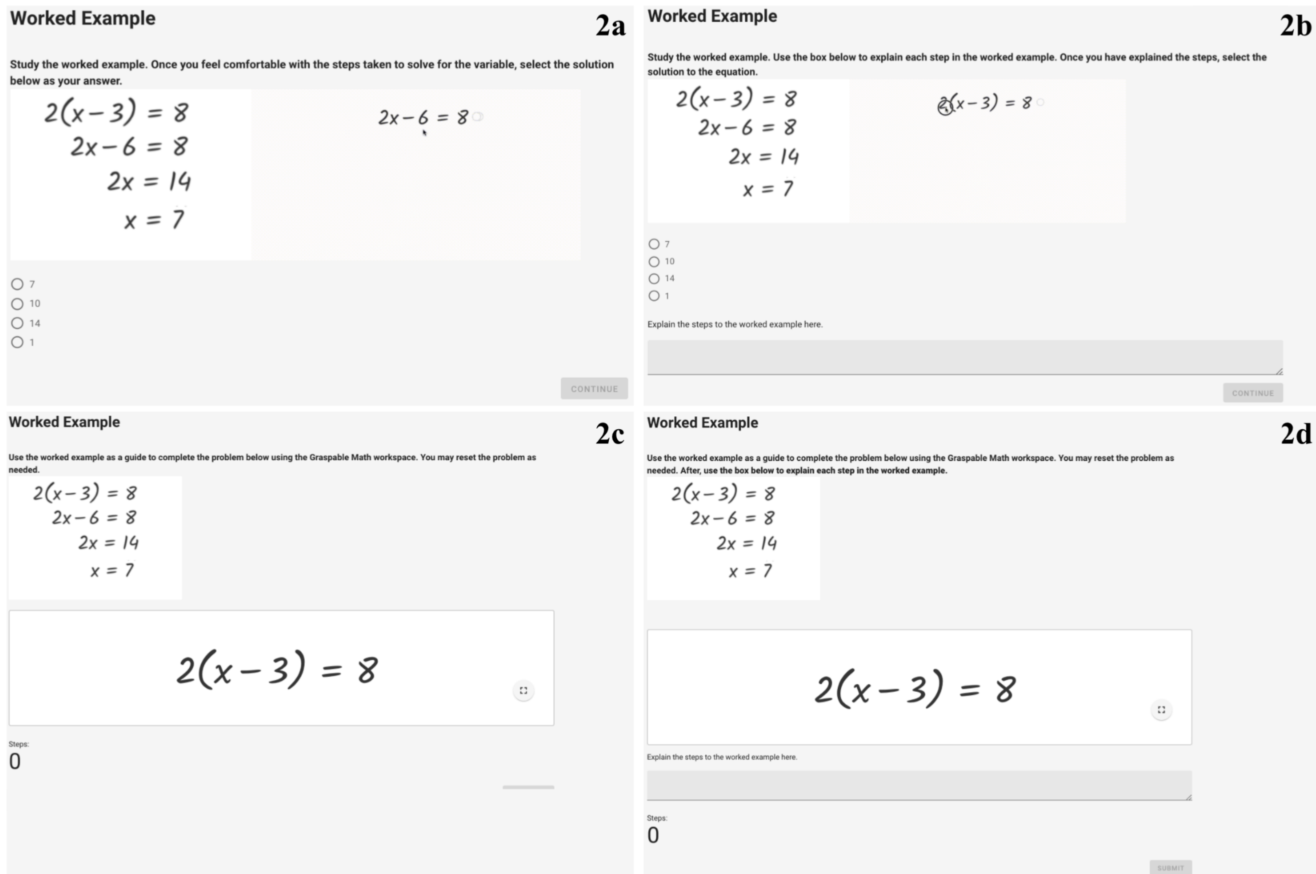
View-and-Explain Condition. Students in the *view-and-explain* condition were prompted to study the same worked examples presented in the view condition. Additionally, they were prompted to provide a typed explanation for each step of the worked example. Specifically, students were prompted to "*Study the worked example. Use the box below to explain each step in the worked example. Once you have explained the steps, select the solution to the equation*." Beneath the worked example, students also saw a free-response box and the following instructions: "*Explain the steps to the worked example here*." (Figure 3b). Like the view condition, I included the multiple choice question to ensure that the students studied the worked examples.

Mirror Condition. For each worked example problem, students in the *mirror* condition saw a static image of a worked example displayed at the top of the screen, and an interactive problem equation presented in the middle of the screen. They were instructed to manipulate the equation using GM in order to match their solution steps with the worked example image. Specifically, students saw the following prompt: "*Use the worked example as a guide to complete the problem below using the Graspable Math workspace. You may reset the problem as needed*." (Figure 3c).

Mirror-and-Explain Condition. In addition to mirroring worked examples, students in the *mirror-and-explain* condition were also prompted to provide a self-explanation of each step taken. Specifically, students saw the following prompt: "*Use the worked example as a guide to complete the problem below using the Graspable Math workspace. You may reset the problem as needed. After, use the box below to explain each step in the worked example*." Beneath the worked example and workspace, students saw a text box instructing, "*Explain the steps to the worked example here*." (Figure 3d).

Figure 3

Worked Examples in the a) View, b) View-and-Explain, c) Mirror, and d) Mirror-and-Explain Conditions.



Note. In the viewing conditions (2a and 2b), students saw a looping video of the worked example transformation process on the right side of the screen. In the explain conditions (2b and 2d), students were prompted to describe the steps in the worked example. In the mirror conditions (2c and 2d), students transformed their expression (e.g., distributing 2 to (x - 3)) to match the worked example.

Cognitive Load Measure

To measure students' cognitive load immediately after completing the intervention, I made minor modifications to an instrument created and validated by Leppink and colleagues (2013) in a series of studies. Notably, whereas most prior instruments have not attempted to distinguish between different types of cognitive load (Sweller et al., 2019), this 10-item instrument captures levels of intrinsic load (items 1, 2, 3; $\alpha > .80$ across studies), extraneous load (items 4, 5, 6; $\alpha \geq .75$ across studies), and germane load (items 7, 8, 9, 10; $\alpha > .80$ across studies). This measure of cognitive load has since been used by Leppink and colleagues (2014) as well as more recently by Tang and colleagues (2019) in a study on the effects of physical tracing during worked example practice on learning. Here, I slightly modified the language used in each item to match the scope of the current study while keeping the 11-point Likert scale consistent with previous work (Appendix C). Students' cognitive load score was calculated as the average of the 10 items. Students' intrinsic, extraneous, and germane load were respectively calculated as the average across the items pertaining to that construct.

Day Three: Posttest

An eight-item posttest with items that mirror the pretest in equation structure assessed students' knowledge after the intervention (Appendix A). Like the pretest, students were instructed to solve for the variable in each equation. Additionally, students completed a brief demographics survey to specify their gender, age, grade level, and prior experience with GM.

Approach to Analysis

Students' pretest, posttest, and demographic data were pulled from Qualtrics and students' behavior and problem-solving data were pulled from GM log files. These data were combined and aggregated at the student-level (i.e., one row per student) for analyses. First, I conducted descriptive statistics to cull cases listwise as needed. Next, for primary analyses, I tested our first three hypotheses using JASP software (JASP Team, 2020; Wagenmakers et al., 2018) to conduct a 2 (Time: Pretest and Posttest) $\times$ 2 (Presentation: View vs Mirror) $\times$ 2 (Self-Explanation: Prompt or No Prompt) repeated measures ANOVA. I conducted the ANOVA using both frequentist and Bayesian methods. In addition to reporting the frequentist statistics, the Bayesian results provide multiple affordances over frequentist analyses such as testing the null hypothesis directly and providing another interpretation of the results (van de Schoot & Depaoli, 2014). For the Bayesian ANOVA, I used the default, non-informative prior specifications in JASP as recommended because I did not have sufficient information to use an informed prior. The default specification uses a JZS (multivariate Cauchy) prior on the effect scales with a default scale of 0.5.

I interpreted the Bayes Factor (BF_{10}) based on common thresholds (Schönbrodt & Wagenmakers, 2018). Specifically, a value smaller than 1, 1/3, and 1/10 provides anecdotal, moderate, and strong evidence for the null hypothesis, respectively. Similarly, a value greater than 1, 3, and 10 provides anecdotal, moderate, and strong evidence for the alternative hypothesis, respectively. This analysis plan allowed us to detect evidence in favor of the experimental or null hypothesis for each of our research questions as follows:

1. The Time $\times$ Presentation effect detected whether students learn more from viewing or mirroring worked examples to answer our first research question.

2. The Time ✕ Self-Explanation effect detected whether students learn more from worked examples with or without self-explanation prompts to answer our second research question.
3. Last, the Time ✕ Presentation ✕ Self-Explanation effect indicated whether there was an interaction between worked example presentations and self-explanation prompts on learning in order to answer our third research question.

Finally, to explore the impact of each worked example condition on students' cognitive load levels, I conducted a series of 2 (Presentation) ✕ 2 (Self-Explanation) ANOVAs comparing students' self-reported cognitive load levels by condition. First, students' average cognitive load score was the outcome variable, followed by students' average scores for intrinsic load (items 1, 2, 3), extraneous load (items 4, 5, 6) and germane load (7, 8, 9, 10). These analyses were interpreted based on the same guidelines reported above.

Results

Preliminary Analyses

Table 1

Descriptive Statistics of Each Variable (N=64)

	Mean	SD	Min–Max	Skewness	Kurtosis
Pretest Score	.39	.25	0-.75	-.16	-1.29
Posttest Score	.60	.30	0-1	-.5	-.88
Cognitive Load Score	3.40	1.75	0-8.6	.19	-.01
Intrinsic Load	2.47	2.54	0-10	1.2	.96
Extraneous Load	2.66	1.82	0-8	.45	-.18
Germane Load	5.05	2.90	0-10	.04	-1.02

Table 1 presents descriptive statistics, including the range and distribution, of focal variables for the full sample. All variables reflect close to normal distributions. Importantly, no students scored at ceiling on the pretest. While 10 (16%) students scored at floor on the pretest and two (3%) students scored at floor on the posttest, students' scores suggest that these items were difficult but appropriate for participants in the sample, leaving room for improvement from pretest to posttest. These students were included in analyses due to the small sample size and preliminary nature of these results. Additionally, a one-way ANOVA did not detect any significant differences in pretest score by condition ($F[3,60]= 0.932$, $p = .431$, $\eta^2 = .045$).

Table 2

Average (SD) of Focal Variables by Condition

	View (*n* = 16)	View and Explain (*n* = 14)	Mirror (*n* = 16)	Mirror and Explain (*n* = 18)
Pretest Score	.42 (.25)	.45 (.25)	.31 (.23)	.38 (.27)
Posttest Score	.52 (.31)	.70 (.23)	.52 (.34)	.66 (.27)
Cognitive Load	3.41 (1.69)	2.65 (1.63)	3.46 (1.52)	3.83 (2.04)
Intrinsic Load	2.41 (2.03)	1.13 (1.43)	2.17 (1.99)	3.70 (3.43)
Extraneous Load	2.68 (1.47)	2.36 (1.39)	2.72 (1.73)	2.80 (2.48)
Germane Load	5.07 (2.44)	4.53 (3.94)	5.56 (2.61)	4.97 (2.86)

Note: SD = standard deviation; mins = minutes

Below, Table 3 presents the correlations between each focal variable. As expected, students' pretest and posttest scores were positively correlated with one another. Also as expected, students' overall cognitive load was positively correlated with each construct within the scale (i.e., intrinsic, extraneous, and germane load). However, the subscales were low-to-moderately correlated with one another, motivating the need to analyze the subscales individually in addition to students' aggregate cognitive load score.

Table 3

Correlations between each variable

	Pretest Score	Posttest Score	Cognitive Load Score	Intrinsic Load	Extraneous Load	Germane Load
Pretest Score	–					
Posttest Score	.667***	–				
Cognitive Load Score	-.226	-.142	–			
Intrinsic Load	-.191	-.138	.746***	–		
Extraneous Load	-.021	$-2.711 * 10^{-4}$	.639***	.384**	–	
Germane Load	-.175	-.031	.731***	.279*	.140	–

Note: Values represent Pearson's r for the correlation coefficient.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

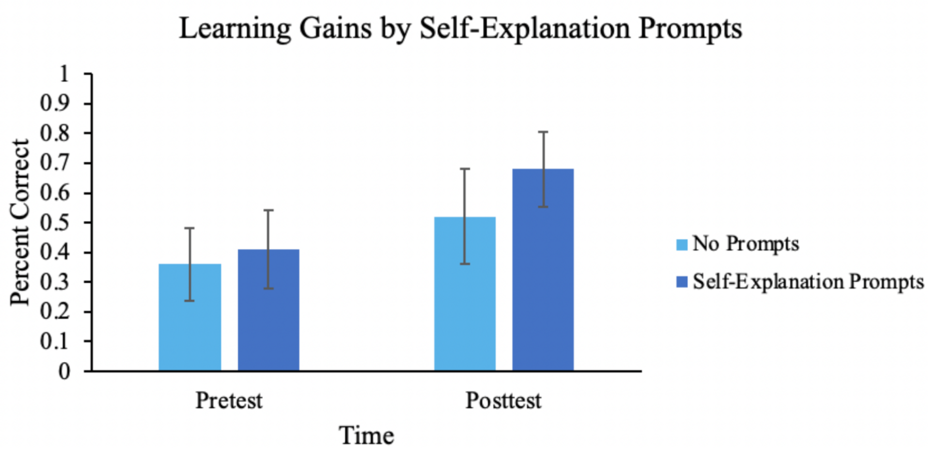
Primary Analysis

A 2 (Time: Pretest vs Posttest) ✕ 2 (Presentation: View vs Mirror) ✕ 2 (Self-Explanation: Prompt or No Prompt) repeated measures ANOVA revealed a main effect of time (F[1,60]= 56.93, $p < .001$, $\eta_p^2 = .487$) showing that, in general, students improved from pretest ($M = .39$, $SD = .25$) to posttest ($M = .60$, $SD = .30$). The Bayesian ANOVA provided strong evidence in support of the main effect of time ($BF_{10} = 1.907 * 10^7$). There was no main effect of mirroring vs. viewing worked examples (F[1,60]= 0.82, $p = .37$, $\eta_p^2 = .013$), with anecdotal evidence in support of this null effect ($BF_{10} = .377$). There was a marginal effect of self-explanation prompts (F[1,60]= 2.80, $p = .10$, $\eta^2 = .045$) but the Bayesian ANOVA provided anecdotal evidence in support of the null hypothesis (i.e., suggesting no effect of self-explanation prompts; $BF_{10} = .872$). There was also a marginal Time ✕ Self-explanation prompt interaction

($F[1,60]= 3.74$, $p = .06$, $\eta_p^2 = .059$), with strong evidence in support of this interaction ($BF_{10} = 1.876*10^7$). Post hoc comparisons with Bonferroni corrections revealed that all students performed comparably at pretest and significantly improved from pretest to posttest ($p < .001$); however, this effect was larger for students in the self-explanation prompt conditions. Specifically, students who received self-explanation prompts demonstrated larger growth (pretest: $M = .41$, $SD = .26$; posttest: $M = .68$, $SD = .25$; $p < .001$) who did not receive self-explanation prompts (pretest: $M = .36$, $SD = .24$; posttest: $M = .52$, $SD =.32$; $p = .001$; Figure 4).

Figure 4

Time ✕ Self-Explanation Prompt Effect on Student Learning Through Worked Examples



Note: Error bars represent the standard deviation.

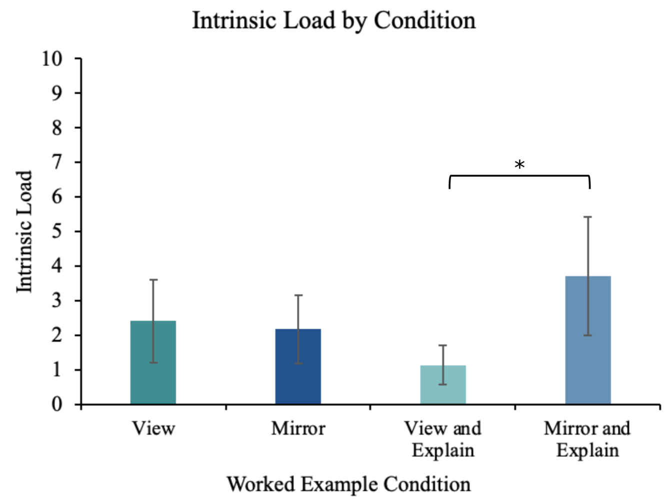
Effects of Worked Example Condition on Cognitive Load

To answer our fourth research question investigating whether students reported different levels of cognitive load by condition, I ran a series of 2 (Presentation) ✕ 2 (Self-Explanation) ANOVAs and repeated them as Bayesian ANOVAs. First, an ANOVA revealed no main effect of presentation ($F[1,58]= 1.887$, $p = .175$, $\eta_p^2 = .032$) or self-explanation prompts on students' overall cognitive load ($F[1,58]= 1.181$, $p = .672$, $\eta_p^2 = .003$), with anecdotal-to-moderate evidence in support of these null findings (BF_{10} = .522 and BF_{10} = .262). Next, an ANOVA revealed no main effect of presentation (F[1,57]= 0.244, p = .623, η_p^2 = .004) or self-explanation prompts on students' extraneous cognitive load (F[1,57]= 0.062, p = .804, η_p^2 = .001), with moderate evidence in support of these null findings (BF_{10} = .286 and BF_{10} = .264). Similarly, there was no main effect of presentation (F[1,57]= 0.369, p = .546, η_p^2 = .006) or self-explanation prompts on students' germane load (F[1,57]= 0.555, p = .459, η_p^2 = .010), with moderate evidence in support of these null findings (BF_{10} = .294 and BF_{10} = .319).

However, an ANOVA did detect a marginally significant effect of presentation (F[1,58]= 3.508, p = .066, η_p^2 = .057) on students' intrinsic load, with anecdotal-to-no evidence supporting this result (BF_{10} = 0.949). There was no main effect of self-explanation prompts on students' intrinsic load (F[1,58]= 0.042, p = .839, η_p^2 < .001), with moderate evidence supporting this null result (BF_{10} = 0.301). Further, there was a Presentation ✕ Self-Explanation interaction (F[1,58]= 5.086, p = .028, η_p^2 = .081), with the Bayesian ANOVA detecting anecdotal evidence to support this model (BF_{10} = 0.662). Post hoc comparisons with Bonferroni corrections showed that students in the mirror-and-explain condition reported significantly higher levels of intrinsic load (*M* = 3.70, *SD* = 3.43) than students in the view-and-explain condition (*M* = 1.13, *SD* = 1.43; *p* = .037). No other pairwise comparisons were significant (*p*s > .05; Figure 5).

Figure 5

Students' Self-Reported Intrinsic Cognitive Load by Condition



Note: Error bars represent the standard deviation.
* $p < .05$

Discussion

This study investigates how different features of worked examples impact student learning from a brief intervention. From the preliminary results reported, four notable preliminary findings have emerged. First, all students, on average, improved from pretest to posttest, demonstrating learning gains across the conditions. Second, there was no evidence to suggest that viewing or mirroring worked examples was more effective for learning. Third, preliminary evidence suggests that students learned more from studying worked examples with self-explanation prompts than without self-explanation prompts. And fourth, students in the mirror-and-explain condition reported higher levels of intrinsic load than students in the view-and-explain condition. I expand on each of these findings and discuss interpretations in the following sections.

The Worked Example Effect

The preliminary evidence revealed that, on average, students across each of the four conditions improved approximately 20% from pretest to posttest after participating in the intervention with worked examples. This finding is aligned with prior research on the worked example effect showing that worked example practice increases learning beyond solving practice problems alone (Booth et al., 2013; Carroll, 1994; Foster et al., 2018). Since the worked example effect is well-supported to date, I did not include a control condition without worked examples. Instead, these preliminary findings suggest that the worked example effect may be very robust. Specifically, perhaps the format of worked examples does not matter as much as the act of engaging in worked example practice itself. Finding ubiquitous learning gains is also aligned with our previous work (Smith et al., 2022) in which we found that students experienced learning gains across conditions after studying one of six different presentations of worked examples.

Finding preliminary evidence of learning gains across all participants is a valuable contribution to the literature. Since limited research has tested the effects of interactive worked example formats in online settings for algebra, these preliminary results add to the literature by indicating that a variety of worked example formats may be effective for supporting learning in online settings. In particular, the ubiquitous learning gains suggest that worked examples are an effective instructional support, and that the variation in the design and presentation of worked examples may not significantly impact learning in online contexts; providing implications for researchers, educators, and content developers who design worked examples for online contexts. In addition to motivating full data collection for the current study, these preliminary findings invite future research to explore how pervasive the effects of different worked example formats are in online settings.

Viewing vs. Mirroring Worked Examples

In testing whether mirroring worked examples may be more effective than simply viewing them, this project is timely: cognitive load theory has been the predominant explanation for the effectiveness of worked examples for thirty years (e.g., Sweller, 1988, Chandler & Sweller, 1991) while more recently, theories of embodied cognition have been gaining traction across multiple areas of psychological science (e.g., Abrahamson et al., 2020; Nathan, 2021). The preliminary findings suggest that students demonstrated comparable learning gains regardless of whether they viewed or mirrored worked examples, suggesting that each format may have its advantages for learning. Further, preliminary results found differences in students' self-reported levels of intrinsic load between those who viewed-and-explained worked examples and those who mirrored-and-explained worked examples. These findings suggest that even if differences in worked example presentation do not impact students' overall learning, different worked example presentations may still affect students' cognitive processes at a more granular level.

Looking ahead, findings from a larger sample on viewing vs. mirroring worked examples will contribute to our understanding of cognition by suggesting how principles of cognitive load theory and embodied cognition influence learning. For instance, if viewing worked examples leads to higher learning gains, these findings will provide further support for using cognitive load theory to explain underlying processes of learning. However, if students who mirror worked examples experience higher learning gains, these findings will suggest that theories of embodied cognition may also be involved in the process of learning. If the latter, these findings will prompt further research to tease apart the influence of cognitive load and the influence of embodied experiences on learning to respectively advance our understanding of these two cognitive theories and how aspects of each theory may be applied together to support learning.

Additionally, this study provides just one example of how student interaction and embodiment may be incorporated in online worked examples. Multiple educational platforms and technologies leverage students' perceptual processes and actions to develop mathematical reasoning (e.g., The Hidden Village: Nathan & Walkington, 2017; the Mathematical Imagery Trainer: Abrahamson & Trninic, 2015; Geogebra). These platforms and technologies could consider designing different formats of embodied worked examples specific to the affordances of their own design. For instance, The Hidden Village may leverage their AR technology to study how students develop mathematical reasoning from worked examples that involve no movement, partial-movement, or full-body movement with provide perceptual feedback from the system. As educational technologies continue to develop, so do the possibilities for incorporating principles of embodied design into instructional support like worked examples.

The Impact of Self-Explanation Prompts

This study, and the preliminary evidence, supports prior research on the use of self-explanations for learning in worked examples. I predicted that students who received self-explanation prompts in their worked examples would improve more from pretest to posttest than their counterparts and the preliminary evidence indicates emerging findings in that direction. These preliminary results conceptually replicate prior work showing that worked examples with self-explanation prompts are more effective for learning than worked examples alone, providing more evidence to support theory on the role of self-explanations in learning from worked examples (Aleven & Koedinger, 2002; Berthold et al., 2009; Chi et al., 1989; Nokes-Malach et al., 2013; Renkl, 2002; 2014).

Further, prior to this study, no prior research had investigated how self-explanations interacted with different worked example formats, particularly those that involve dynamically interacting with worked examples on-screen. The preliminary findings from this study suggest that self-explanations are a robust and powerful aid in learning from worked examples, regardless of the worked example format. Looking ahead, findings from a larger sample will provide practical implications for designing worked examples for online settings by identifying when worked examples with self-explanations are effective.

The Impact of Worked Examples on Cognitive Load

Beyond comparing learning across conditions, I analyzed students' self-reported levels of cognitive load to delineate how studying different worked example formats may impact student learning at a more granular level. I predicted that the mirroring presentations of worked examples would not increase students' extraneous cognitive load but may increase germane cognitive load by providing additional cues in the worked examples. With the current sample, I found no differences in students' overall self-reported cognitive load by condition. Looking into the three types of cognitive load, I found no differences in students' self-reported levels of extraneous or germane load. However, contrary to my predictions I found that students in the mirror-and-explain condition reported significantly higher levels of intrinsic load than students in the view-and-explain condition.

At first glance, this finding might suggest that the task demands of mirroring and explaining worked examples increases cognitive load more than just viewing and explaining worked examples. However, the intrinsic load items on the survey adapted from Leppink and colleagues (2013) address the complexity of the subject matter itself (i.e., algebraic equations) rather than the complexity of the worked examples (extraneous load) or the overall activity (germane load). This finding indicates that students differed in their perceived complexity of the subject matter rather than the cognitive strain experienced by studying the worked examples or participating in the overall instructional activity. Given that I did not expect to find differences in students' intrinsic load by condition, this result raises speculation. On the one hand, students' pretest scores were not significantly correlated with intrinsic load. It is possible that other variables not included in this study may explain this finding or perhaps the demands of mirroring and explaining worked examples increases the perceived complexity of simplifying equations. On the other hand, this sample was small enough that no significant differences were detected in pretest performance across conditions although anecdotally, students in the mirror-and-explain condition did average approximately 7% lower on pretest than students in the view-and-explain condition. This practical difference in pretest performance suggests that the difference in reported intrinsic load may be attributed to differences in students' prior knowledge by condition rather than a reflection of the worked example designs.

I anticipate that data analysis with a full sample may reveal differences in cognitive load across conditions; thereby, teasing apart some of the mechanisms through which worked examples influence learning. For example, I anticipate that students with no prior experience using GM may report higher levels of extraneous and germane load as they manage cognitive demands associated with using, and learning from, a new technology tool. Results from the full sample will shed light on how worked examples with different features may impact students' cognitive processes to inform the design of worked examples for online environments.

Limitations and Future Directions

The study and reported results have multiple limitations that invite future directions. Namely, the preliminary results reported are from a small sample, rendering the results largely inconclusive. In response to the small sample size, I opted to conduct data analysis without accounting for individual differences that may have impacted treatment effects such as students' prior knowledge, prior experience with GM, or time spent on the worked examples to control for potential differences from condition to condition (e.g., Zhou et al., 2015).

Looking ahead, I plan to conduct a full round of data collection and conduct data analysis that will more accurately estimate treatment effects by controlling for extraneous factors. For example, in the current analyses, I did not control for students' time spent on worked examples due to the small sample size. With the full sample, I will include students' average time spent on the four worked example problems as a covariate in the statistical models to control for any remaining differences between conditions. Including average time spent on worked examples as a covariate will ensure that the estimated condition effects are independent of time spent on worked examples. Similarly, I will include students' performance at pretest as a covariate in the final analyses to account for any potential effects of students' prior knowledge of simplifying equations and estimate the condition effects independent of prior knowledge. Additionally, since prior experience using GM may prime students to perceive mathematical symbols as tangible objects, it might subsequently impact how students behave and perform in this study. Therefore, I will also include prior experience with GM as a binary covariate in our analyses with the full sample to estimate the condition effects independent of students' prior experience using GM.

Another prominent limitation is the ability to isolate the effect of students' actions (i.e., interacting with GM to mirror worked examples) on learning and cognitive load. For example, students in the viewing condition may have copied worked examples on paper while completing the intervention. If that happened, those students may have benefited from "mirroring" the worked examples on paper similar to the students who mirrored the worked examples on-screen.

Conversely, we did not enforce standards for the self-explanations so some students may have provided brief explanations with little-to-no rationale. Future work should seek to isolate the effects of viewing vs mirroring and self-explanation prompts to draw causal influence about how students' interactions with worked examples impacts learning.

Further, this study invites multiple avenues of future research that coincide with learning analytics and the growing synergy between learning analytics and embodied design (Abrahamson et al., 2021). First, it may be worthwhile to analyze students' log data from GM during the intervention to test whether students' behavior during the intervention varies depending on the worked examples they study. Similarly, I did not analyze students' self-explanations in this manuscript. Looking ahead, I plan to leverage natural language processing techniques to explore patterns of speech detected in students' self-explanations over time and across conditions. For instance, under the assumption that mirroring worked examples in GM allows students to treat math symbols as objects—moving and manipulating them on-screen—I would predict that students in the mirror-and-explain condition would use more dynamic language to describe the worked examples than their counterparts in the view-and-explain condition. Additionally, rather than an online experiment, we may see more mechanisms of students' learning processes by conducting this study in a one-on-one interview format to record students' speech and gestures. I would predict that if students were asked to verbally explain the steps taken in the worked examples to a researcher, we would see different patterns of speech as well as gestures between students in the mirror-and-explain and view-and-explain conditions. Together, these avenues of future work demonstrate the complexity of example-based learning and the benefits of applying an interdisciplinary approach to the study of example-based learning.

Implications and Conclusions

This study contributes preliminary evidence to advance theories of cognition and learning; simultaneously, it provides recommendations to researchers, teachers, and content developers for how worked examples can and should be implemented in online learning environments to support student learning in algebra. First, this study demonstrates one way that researchers, teachers, and content developers may draw upon theories of cognitive load as well as embodied cognition to design worked examples for online contexts that leverage affordances unique to educational technologies. Second, the preliminary results demonstrate that the worked example effect may be robust and not contingent on a specific format, suggesting that there is flexibility in how worked examples may be effectively designed for online contexts. Preliminary evidence also suggests that self-explanation prompts contributed to learning beyond studying worked examples alone, suggesting that, in online settings across different formats, teachers and content developers can consider pairing worked examples with self-explanation prompts. However, these conclusions are inconclusive without further evidence from a larger sample to increase power.

Looking ahead, finding any reliable differences in learning gains between two or more conditions with a larger sample will inform cognitive and learning theory as well as offer recommendations to design worked examples for online settings. As more online learning platforms become available for algebra practice (e.g., Graspable Math, Weitnauer et al., 2016; ASSISTments, Heffernan & Heffernan, 2014; Cognitive Tutor; Ritter et al., 2007), novel research and cognitive theory should continue to help shape the design of instructional support in these learning environments. To that end, this study provides an avenue to investigate how worked examples may be designed to balance cognitive demand and support in online learning environments. Beyond the interpretations provided for the preliminary results reported here, I anticipate that findings from a larger sample will provide theoretical contributions as well as implications for future research, technology design, and educational practice.

References

Abrahamson, D., Nathan, M. J., Williams-Pierce, C., Walkington, C., Ottmar, E. R., Soto, H., & Alibali, M. W. (2020). The future of embodied design for mathematics teaching and learning. *Frontiers in Education, 5,* 1-29. https://doi.org/10.3389/feduc.2020.00147

Abrahamson, D., and Trninic, D. (2015). Bringing forth mathematical concepts: signifying sensorimotor enactment in fields of promoted action. *ZDM Math. Educ.* 47, 295–306. doi: 10.1007/s11858-014-0620-0

Abrahamson, D., Worsley, M., Pardos, Z. A., & Ou, L. (2021). Learning analytics of embodied design: Enhancing synergy. *International Journal of Child-Computer Interaction*, 100409.

Adkins, M., & Noyes, A. (2018). Do advanced mathematics skills predict success in biology and chemistry degrees? *International Journal of Science and Mathematics Education*, *16*(3), 487-502.

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer‐based cognitive tutor. *Cognitive Science*, *26*(2), 147-179.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice Hall.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617-645.

Barbieri, C., & Booth, J. L. (2016). Support for struggling students in algebra: Contributions of incorrect worked examples. *Learning and Individual Differences*, *48*, 36-44.

Barbieri, C. A., & Booth, J. L. (2020). Mistakes on display: Incorrect examples refine equation solving and algebraic feature knowledge. *Applied Cognitive Psychology, 34*, 862–878. https://doi.org/10.1002/acp.3663

Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science, 37,* 345–363.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24-34.

Booth, J. L., Oyer, M. H., Paré-Blagoev, E. J., Elliot, A. J., Barbieri, C., Augustine, A., & Koedinger, K. R. (2015). Learning algebra by example in real-world classrooms. *Journal of Research on Educational Effectiveness*, *8*(4), 530-551.

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*(4), 539.

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, *86*(3), 360.

Chan, J. Y.-C., Lee, J.-E., Mason, C. A., Sawrey, K., & Ottmar, E. (2021). From Here to There! A dynamic algebraic notation system improves understanding of equivalence in middle-school students. *Journal of Educational Psychology.* Advance online publication. https://doi.org/10.1037/edu0000596

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293-332.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, *10*(2), 151-170.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439-477.

Chi, M. T., & VanLehn, K. A. (1991). The content of physics self-explanations. *The Journal of the Learning Sciences*, *1*(1), 69-105.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047-1058.

Cooper, G., & Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79,* 347–362.

Dörfler, W. (2003). Mathematics and mathematics education: Content and people, relation and difference. *Educational Studies in Mathematics*, *54*(2), 147-170.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58.

Faulkenberry, T. J., Ly, A., & Wagenmakers, E. J. (2020). Bayesian inference in numerical cognition: A tutorial using JASP. *Journal of Numerical Cognition*, *6*(2), 231-259.

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(3), 319–325.

Foster, N. L., Rawson, K. A., & Dunlosky, J. (2018). Self-regulated learning of principle-based concepts: Do students prefer worked examples, faded examples, or problem solving?. *Learning and Instruction*, *55*, 124-138.

Ginns, P., Hu, F. T., Byrne, E., & Bobis, J. (2016). Learning by tracing worked examples. *Applied Cognitive Psychology*, *30*(2), 160-169.

Ginns, P., Hu, F. T., & Bobis, J. (2020). Tracing enhances problem‐solving transfer, but without effects on intrinsic or extraneous cognitive load. *Applied Cognitive Psychology*, *34*(6), 1522-1529.

Goldstone, R., Marghetis, T., Weitnauer, E., Ottmar, E., & Landy, D. (2017). Adapting perception, action, and technology for mathematical reasoning. *Current Directions in Psychological Science, 26*(5), 434-441. DOI: https://doi.org/10.1177/0963721417704888

Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior*, *25*(2), 267-274.

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, *18*(1), 54-65.

Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & Van Gog, T. (2019). Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology*, *111*(1), 45.

Hu, F. T., Ginns, P., & Bobis, J. (2015). Getting the point: Tracing worked examples enhances learning. *Learning and Instruction, 35*, 85–93. https://doi.org/10.1016/j.learninstruc.2014.10.002

Hulse, T., Daigle, M., Manzo, D., Braith, L., Harrison, A., & Ottmar, E. (2019). From here to there! Elementary: a game-based approach to developing number sense and early algebraic understanding. *Educational Technology Research and Development*, *67*(2), 423-441.

JASP Team (2020). JASP (Version 0.14.1)[Computer software].

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., ... & Velez, E. D. V. (2015). The Condition of Education 2015. NCES 2015-144. *National Center for Education Statistics*.

Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *The Quarterly Journal of Experimental Psychology*, *63*(10), 1953-1968.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge University Press.

Lee, J. (2013). College for all: Gaps between desirable and actual P–12 math achievement trajectories for college readiness. *Educational Researcher, 42*(2), 78–88.

Leppink, J., Paas, F., Van Gog, T., Van Der Vleuten, C. P. M., & Van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32–42.

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior research methods*, *45*(4), 1058-1072.

Margulieux, L. E., Morrison, B. B., Guzdial, M., & Catrambone, R. (2016). Training learners to self-explain: Designing instructions and examples to improve problem solving. In Looi, C. K., Polman, J. L., Cress, U., and Reimann, P. (Eds.). *Transforming Learning,* *Empowering Learners: The International Conference of the Learning Sciences (ICLS)* 2016, Volume 1. Singapore: International Society of the Learning Sciences. https://www.isls.org/icls/2016/docs/ICLS2016_Volume_1_30June2016.pdf

Mayer, R. E. (2020). Advances in designing instruction based on examples. *Applied Cognitive Psychology*, *34*(4), 912-915.

McLaren, B. M., van Gog, T., Ganoe, C., Karabinos, M., & Yaron, D. (2016). The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in computer-based learning environments. *Computers in Human Behavior*, *55*, 87-99.

Nathan, M. J. (2014). Grounded mathematical reasoning. In L. Shapiro (Ed.). *The Routledge handbook of embodied cognition* (pp. 171-183). Albingdon, UK: Routledge.

Nathan, M. J. (2021). *Foundations of embodied learning: A paradigm for education*. Routledge.

Nathan, M. J., and Walkington, C. (2017). Grounded and embodied mathematical cognition: promoting mathematical insight and proof using action and language. *Cogn. Res. Principles Implicat.* 2:9.

Nokes-Malach, T. J., VanLehn, K., Belenky, D. M., Lichtenstein, M., & Cox, G. (2013). Coordinating principles and examples through analogy and self-explanation. *European Journal of Psychology of Education*, *28*(4), 1237-1263.

Ottmar, E. R., Landy, D., Weitnauer, E., & Goldstone, R. (2015). "Graspable mathematics: using perceptual learning technology to discover algebraic notation," in *Integrating Touch-enabled and Mobile Devices into Contemporary Mathematics Education*.

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*, 27–45. https://doi.org/10.1007/s10648-011-9179-2

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* London: National Research and Development Centre for Adult Literacy and Numeracy.

Reed, S. K., Corbett, A., Hoffman, B., Wagner, A., & McLaren, B. (2013). Effect of worked examples and Cognitive Tutor training on constructing equations. *Instructional Science, 41*(1), 1-24.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*(5), 529-556.

Renkl, A. (2014). Toward an instructionally oriented theory of example‐based learning. *Cognitive Science*, *38*(1), 1-37.

Renkl, A. (2017). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd edn., pp. 325–348). New York, NY: Routledge.

Renkl, A., & Eitel, A. (2019). Self-explaining: Learning about principles and their application. *Cambridge handbook of cognition and education*, 528-549.

Retnowati, E., Ayres, P., & Sweller, J. (2010). Worked example effects in individual and group work settings. *Educational Psychology*, *30*(3), 349-367.

Rittle-Johnson, B., & Loehr, A. M. (2017). Eliciting explanations: Constraints on when self-explanation aids learning. *Psychonomic Bulletin & Review*, *24*(5), 1501-1510.

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561.

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction*, *19*(2), 185-199.

Salden, R. J., Koedinger, K. R., Renkl, A., Aleven, V., & McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educational Psychology Review*, *22*(4), 379-392.

Schalk, L., Roelle, J., Saalbach, H., Berthold, K., Stern, E., & Renkl, A. (2020). Providing worked examples for learning multiple principles. *Applied Cognitive Psychology*, *34*(4), 813-824.

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128-142.

Smith, H., Closser, A. H., Ottmar, E. R., & Chan, J. Y. C. (2022). The impact of algebra worked example presentations on student learning. *Applied Cognitive Psychology,* 1-15. https://doi.org/10.1002/acp.3925

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 81*, 457-466.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295-312.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*.

Sweller, J. & Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*.

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261-292.

Tang, M., Ginns, P., & Jacobson, M. J. (2019). Tracing enhances recall and transfer of knowledge of the water cycle. *Educational Psychology Review*, *31*(2), 439-455.

VanDerHeyden, A. M., & Burns, M. K. (2009). Performance indicators in math: Implications for brief experimental analysis of academic performance. *Journal of Behavioral Education*, *18*(1), 71-91. https://doi.org/10.1007/s10864-009-9081-x

VanLehn, K., Jones, R. M., & Chi, M. T. (1992). A model of the self-explanation effect. *The journal of the Learning Sciences*, *2*(1), 1-59

van Gog, T., Hoogerheide, V., & van Harsel, M. (2020). The role of mental effort in fostering self-regulated learning with problem-solving tasks. *Educational Psychology Review*, 1-18.

van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*, 212–218.

van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*(2), 155-174.

van Harsel, M., Hoogerheide, V., Verkoeijen, P., & van Gog, T. (2019). Effects of different sequences of examples and problems on motivation and learning. *Contemporary Educational Psychology*, 58, 260-275.

van Harsel, M., Hoogerheide, V., Verkoeijen, P., & van Gog, T. (2020). Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences. *Applied Cognitive Psychology, 34,* 793–812. https://doi.org/10.1002/acp.3649 van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *The European Health Psychologist*, *16*(2), 75-84.

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... & Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58-76.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction, 7,* 1–39.

Weitnauer, E., Landy, D., & Ottmar, E. (2016, December). Graspable math: Towards dynamic algebra notations that support learners better than paper. In *2016 Future Technologies Conference (FTC)* (pp. 406-414). IEEE.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9(*4), 625–636.

Yeo, L. M., & Tzeng, Y. T. (2020). Cognitive effect of tracing gesture in the learning from mathematics worked examples. *International Journal of Science and Mathematics Education*, *18*(4), 733-751.

Zhi, R., Price, T. W., Marwan, S., Milliken, A., Barnes, T., & Chi, M. (2019, February). Exploring the impact of worked examples in a novice programming environment. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 98-104).

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology*, 94, 660–668.

Zhou, G., Price, T. W., Lynch, C. F., Barnes, T., & Chi, M. (2015). The Impact of Granularity on Worked Examples and Problem Solving. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society, CogSci 2015*, Pasadena, California, USA, pp. 2817-2822.

Appendix A. Pretest and Posttest Items.

Pretest Items	Source	Posttest Items	Source
$8(2x + 9) = 56$	Engage NY	$11(x + 10) = 132$	Engage NY
$-(x - 5) + 2 - x = 3$	Project Utah	$-(4x - 10) + 4 - 4x = 6$	Author
$5 - 4(2b - 5) + 3b = 15$	Project Utah	$30 - 4(b - 5) + 1b = 20$	Author
$10 = 3(x - 2) - 2(5x - 1)$	Project Utah	$20 = 3(2x-2) - 2(5x-1)$	Author
$3(2x - 14) + x = 15 - (-9x - 5)$	Engage NY	$6(4x-28) + 2x = 30 - (-18x - 10)$	Author
$-4x - 2(8x + 1) = -(-2x - 10)$	Engage NY	$-6x-4(3x+2) = -(-1x-2)$	Author
$5(y - 12) = 3(y - 12) + 20$	Author	$2(y - 4) = (y - 4) + 6$	Author
$3(h + 2) + 4(h + 2) = 35$	Author	$2(h + 1) + 4(h + 1) = 12$	Author

Note. Engage NY and Project Utah are open-source curricula. Problems created by the author were adapted from open source content to match equation structures between the pretest and posttest items. Items 2, 3, 5, and 6 are transfer items.

Appendix B. Algebraic equations used in worked examples across all conditions and the paired problem completed by students after each worked example.

Worked Examples	Paired Practice Problem
$2(x\text{-}3) = 8$ $2x\text{-}6 = 8$ $2x = 14$ $x = 7$	$-3(y - 4) = 18$
$2(t\text{-}1) + 3(t\text{-}1) = 10$ $2t - 2 + 3t - 3 = 10$ $5t - 5 = 10$ $5t = 15$ $t = 5$	$3(t\text{-}1) + 3(t\text{-}1) = 30$
$5(y+1) = 3(y+1) + 8$ $5y + 5 = 3y + 3 + 8$ $5y + 5 = 3y + 11$ $2y + 5 = 11$ $2y = 6$ $y = 3$	$5(m + 4) = 2(m + 4) + 15$
$9 = 5(m + 2) + 4(m + 2)$ $9 = 5m + 10 + 4m + 8$ $9 = 9m + 18$ $-9 = 9m$ $-1 = m$	$9 = 3(y + 5) + 6(y +5)$

Note. The worked example derivations are those used by Rittle-Johnson and Star (2007).

Appendix C. Cognitive load instrument adapted from Leppink and colleagues (2013) for the measurement of intrinsic load (items 1, 2, 3), extraneous load (items 4, 5, 6), and germane load (items 7, 8, 9, 10).

Instructions: All of the following questions refer to the activity that just finished. Please respond to each of the questions on the following scale (0 meaning *not at all the case* and 10 meaning *completely the case*).

[1] The topic covered in the activity was very complex.

[2] The activity covered equations that I perceived as very complex.

[3] The activity covered concepts that I perceived as very complex.

[4] The worked examples during the activity were very unclear.

[5] The worked examples were, in terms of learning, very ineffective.

[6] The worked examples were full of unclear language.

[7] The activity really enhanced my understanding of the topic(s) covered.

[8] The activity really enhanced my knowledge and understanding of solving equations.

[9] The activity really enhanced my understanding of the equation-solving strategies covered.

[10] The activity really enhanced my understanding of equation-solving.

Chapter 5. Discussion

This chapter summarizes key takeaways from the research studies presented in this dissertation as well as interpretations drawn from the cumulative findings. I address prominent limitations and close with a summary of related ongoing and future work to continue this research program.

Summary and Interpretations of Findings

The studies presented describe: 1) how perceptual cues impact students' performance on simplifying arithmetic problems in an online tutoring system (Chapter 2), 2) the relation between students' prior knowledge, inhibitory control, and perceptual cues (Chapter 3), and 3) how embodied features may be integrated into online worked examples and impact students' learning with a dynamic notation tool (Chapter 4). First, the results presented in Chapter 2 conceptually replicate prior research on perceptual learning in an authentic learning environment by demonstrating that physical spacing in arithmetic expressions impacts students' performance in an online learning environment. Second, the study presented in Chapter 3 extends this work by bridging cognitive and developmental theories. The results showed that spacing in arithmetic problems also impacts college students, and may be moderated by students' prior knowledge, although this effect does not seem to be impacted by students' inhibitory control. Third, while the results in Chapter 4 were inconclusive, preliminary evidence contributes to the literature on the worked example effect by theorizing how multiple cognitive theories (i.e., cognitive load and embodiment) may work together to influence example-based learning. Together, these studies demonstrate that perceptual and embodied cues may be integrated in different ways in online learning platforms and that by doing so, they may impact students' performance and learning in different ways (Table 1).

Table 1

Summary of Studies by Platform, Perceptual or Embodied Cue, and Effect

	Platform	Study Population	Content	Perceptual or Embodied Cue	Effect
Study 1	ASSISTments	5-12th grade students	Order of operations	Spatial proximity	Performance: Accuracy
Study 2	Psychopy	College students	Order of operations	Spatial proximity	Performance: Response time
Study 3	Graspable Math Activities	9th grade students	Simplifying equations	Mirroring worked examples	Learning: Inconclusive

Although these research projects involve short experimental studies with subtle manipulations, these findings support the literature on perceptual and embodied learning. Namely, even the smallest details of students' environments can potentially make a difference by leveraging perceptual supports and students' body-based resources to make learning algebra and pre-algebraic concepts easier. What's more, perceptual scaffolds and embodied features can be easily implemented in instructional practices and math content on online platforms. By exploring how we can integrate these supports into online instructional material, we can provide feasible, cost-effective recommendations for researchers, teachers, and content developers that may generalize across different online learning platforms.

Current Limitations Invite Future Directions

Beyond facing recruitment difficulties due to the COVID-19 pandemic, this dissertation research is limited in its scope and consequently, implications for theory and practice. Notably, although conducting online experiments provided clean experimental designs, the presented results neglect the possibility of uncovering more complex outcomes that may have been achieved with a more qualitative approach, such as observing students' speech and gestures in addition to their performance on arithmetic and algebra tasks. Largely, my research to date has focused on general effects of performance and learning found across samples of participants without considering behavioral processes that could influence the impact of perceptual or embodied features of instructional materials on performance or learning. Similarly, these experiments are brief, taking place over the course of roughly one hour. None of this research considers the impact of instructional support with perceptual and embodied support over longer periods of time.

Looking ahead, I aim to advance cognitive theory and methodological practice in the learning sciences and to provide recommendations for technology design and instructional support in online learning environments. I intend to continue developing and scaling my research program to investigate how we can leverage perceptual support and embodied experiences in mathematics education. First, I intend to continue exploring how individual differences, such as inhibitory control, may influence the relationship between perceptual and embodied features of instructional materials and learning in algebra. Second, I intend to explore how instructional support should be implemented in online platforms to effectively help students long term. Third, I plan to develop a second line of exploratory research dedicated to questioning when and how methodologies across the learning sciences should be used in different contexts. Each of these future directions are discussed in more detail in the following sections.

Embracing Complexity with Worked Examples and Instructional Materials

Scheiter (2020) recently commented that the study of example-based learning has advanced in the past twenty years most noticeably by "embracing complexity". Scheiter (2020) reasoned that beyond studying when and why worked examples are beneficial, we also need to consider how individual differences may impact the relation between worked examples and learning. I agree with this call to action and will go further to say that it also seems applicable for worked examples and instructional materials that try to integrate perceptual and embodied features to support learning.

I believe that the study presented in Chapter 4 begins to embrace complexity by investigating not just how students' performance and learning are impacted by exposure to instructional materials but also the underlying mechanisms like cognitive load. In addition to this work, I have also conducted two small studies observing how algebra students (Closser et al., 2022a) and college students (Closser et al., 2022b) perceive the helpfulness of different worked example formats and why. These studies build off the work presented in Smith, Closser, Ottmar, and Chan (2022) which found that algebra students experienced learning gains after exposure to any one of six perceptually different worked examples in instructional practice. By observing students' reactions to the worked examples, I consider these follow-up studies to be foundational steps in embracing complexity in worked examples by analyzing which features students attend to and how that information may inform future design choices in worked examples for algebra.

More broadly, the effectiveness of instructional materials in math are also impacted by students' individual differences. Perhaps most commonly, prior work has demonstrated that students' prior knowledge impacts the effectiveness of interventions with worked examples (e.g., Kalyuga et al., 2001). But beyond prior knowledge, what other individual differences might impact how students learn from worked examples with perceptual support? Recently, Tempelaar and colleagues (2020) demonstrated that example-based learning research may be doing the field a disservice by searching for overall effects rather than accounting for different learning dispositions and behavioral profiles among students. Similarly, Schwaighofer and colleagues (2016) found that components of executive function moderated the worked example effect whereas prior knowledge did not. Together, these studies suggest that there may be other underlying mechanisms behind example-based learning that could have implications for the design and implementation of instructional support. Additionally, to the best of my knowledge, few studies have investigated the role of prior knowledge and individual differences in learning from worked examples that leverage perceptual features. I believe that the study presented in Chapter 3 provides a foundation for uncovering the role that individual differences, such as inhibitory control, play in how students learn from various instructional materials. Looking ahead, I plan to integrate theoretical perspectives to better understand factors, such as individual differences, that possibly influence the effectiveness of worked examples and other instructional materials.

Implementing Long-Term Instructional Support in Online Platforms

Beyond considering how individual differences may impact the way students learn from instructional materials with perceptual support, I also plan to explore how instructional support should be implemented long-term in online platforms to increase the ecological validity of my work and provide more substantiated recommendations for online platforms. The research presented in this dissertation solely includes brief studies involving one to three sessions, preventing me from being able to draw any conclusions about how students may learn from these instructional supports over time. For instance, evidence suggests that using a concreteness fading approach for instructional support best improves student learning over time (e.g., see Fyfe et al., 2014 for a review; Fyfe & Nathan, 2019; Ottmar & Landy, 2017), including fading of worked examples (Miller-Cotto & Auxter, 2019).

To advance this area of research, I intend to examine the effects of concreteness fading with perceptual and embodied scaffolds in online learning environments. I wonder, *how should perceptual scaffolds be implemented in online learning systems to optimize learning over the course of a semester*? Specifically, do students learn more if they start practicing early algebra concepts in problem sets that utilize scaffolds (e.g., through the use of congruent spacing, color, dynamic worked examples) to direct their attention toward the structures of math notation and then slowly remove those supports over time (e.g., by decreasing spacing, removing color, removing dynamic features)? I intend to compare the impact of the concreteness fading approach on student learning gains to their peers who receive no perceptual support or consistent perceptual support over time. To do so, I will use a range of data and mixed methods to study student learning gains over several weeks as well as their behaviors within a problem set. Ultimately, this research will contribute to our understanding of how perceptual features impact student learning and provide recommendations for online learning platforms.

Advancing Methodological Practices in the Learning Sciences

In addition to my primary line of research, I also aim to advance the field's understanding of how methodologies from subdisciplines of the learning sciences may be appropriately used across projects and contexts to provide new insights on learning. In this dissertation research, I primarily applied frequentist analysis methods (including analysis of variance and hierarchical linear models) as well as Bayesian analysis methods to offer multiple interpretations of the results. That said, my research has sparked discussions on the various methodologies used in learning science research and has prompted exploratory work on the applications of qualitative, quantitative, and computational methods in educational research. Looking ahead, I plan to pursue a second line of research focusing on the application of quantitative and computational methods of data analysis in the learning sciences.

My colleagues and I have already made strides in exploring how different methods from educational data mining and learning analytics can be applied to rich datasets from different contexts in the learning sciences beyond large-scale log files. For instance, we used clustering analyses to identify profiles of student behavior during measurement tasks (Harrison, Smith, Botelho, Ottmar, & Arroyo, 2020) then expanded this work to show how machine learning can be applied to relatively small, multimodal datasets from in-person studies and afford more regularization in models (Closser et al., 2021). This work demonstrates the synergy between learning analytics and embodied design (Abrahamson et al., 2021), showing that cross-disciplinary applications of quantitative and computational methods may be beneficial for advancing the field's understanding of how students learn through multimodal tasks. Our findings are directly applicable to possible extensions of the research presented in this dissertation: for instance, machine learning techniques may be applied, as informed by cognitive theories, to further analyze log data (e.g., students' behaviors and performance on the paired practice problems) from the study presented in Chapter 4. As more educational research is conducted in online platforms, it is critical to consider how we can appropriately analyze a variety of data from different sources.

Similarly, I have asked whether we can advance our field by refining the quantitative methods used to draw causal inferences about learning. My colleagues and I have run simulated experiments to determine which and when different modeling approaches are appropriate for analyzing data from online experiments with student-level randomization to account for the naturally nested structure of students in classrooms (Closser et al., in preparation). On the one hand, using student-level randomization in experiments should alleviate any group level differences. On the other hand, my prior work (Chapter 2; Harrison, Smith, Hulse, & Ottmar, 2020) suggests that student-level randomization may not completely offset class- or school-level differences in samples. This discrepancy, along with recent work debating the value of multilevel modeling (e.g., McNeish et al., 2017) suggested that it may be worthwhile, as a field, to reconsider when and how to appropriately analyze data from online experiments in educational research in order to accurately estimate treatment effects. Through this budding line of research, I am motivated to leverage the strengths, and minimize the limitations, of methodologies used in the learning sciences to advance research and provide new insights into learning.

Conclusion

The three projects presented in this dissertation provide a foundational approach to investigating how perceptual scaffolds and embodied features within instructional materials impact students' performance and learning in online settings. I have been excited by this line of research and am eager to continue pursuing it because I believe that investigating how perceptual and embodied learning occur in online settings foots the bill for research in Pasteur's quadrant by balancing principles of basic and applied research (Stokes, 2011). On the one hand, this research contributes to our theoretical understanding of how cognitive processes occur and are influenced by perception and embodied experiences. Further, it presents opportunities to consider how different theoretical perspectives may overlap and have implications for cognitive and learning processes in math. On the other hand, each of these projects provide practical implications for instructors and content developers for online platforms by demonstrating a) how perceptual and embodied scaffolds may be feasibly implemented in online learning environments and b) how students' behavior, performance, and learning outcomes may be impacted by these features. Ideally, this line of research will advance theory while also leading to the creation of guidelines for small, positive changes in the ways that math instructional materials are presented to students in online settings to support math learning.

References

Abrahamson, D., Worsley, M., Pardos, Z. A., & Ou, L. (2021). Learning analytics of embodied design: Enhancing synergy. *International Journal of Child-Computer Interaction*, 100409.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181-214.

Cassotti, M., Agogue, M., Camarda, A., Houde, O., & Borst, G. (2016). Inhibitory control as a core process of creative problem solving and idea generation from childhood to adulthood. In B. Barbot (Ed.), Perspectives on creativity development. *New Directions for Child and Adolescent Development, 151,* 61–72.

Closser, A. H., Erickson, J. A., Smith, H., Varatharaj, A., & Botelho, A. F. (2021, in press). Blending Learning Analytics and Embodied Design to Model Students' Comprehension of Measurement Using Their Actions, Speech, and Gestures. To appear in Abrahamson, D., Worsley, M., Pardos, Z., Ou, L. (Eds.) *International Journal of Child-Computer Interaction (IJCCI) Special Issue on Learning Analytics of Embodied Design: Enhancing Synergy*. https://doi.org/10.1016/j.ijcci.2021.100391

Closser, A. H., Sales, A., & Botelho, A. F. (in preparation). Should we account for classrooms? Analyzing online experimental data with student-level randomization. Manuscript in preparation.

Closser, A. H., Smith, H., & Chan, J. Y. C. (2022, November). *Students' Impressions of Equation-Solving Worked Examples in an Online Environment*. Brief research report to be included in the *Proceedings of the 44thMeeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Nashville, TN.*

Closser, A. H., Chan, J. Y. C, Smith, H., & Ottmar, E. (2022, November). *College Students' Input on The Design of Worked Examples for Online Environments*. Research report to be included in the *Proceedings of the 44thMeeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Nashville, TN.*

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational psychology review*, *26*(1), 9-25.

Fyfe, E. R., & Nathan, M. J. (2019). Making "concreteness fading" more concrete as a theory of instruction for promoting transfer. *Educational Review, 71*, 403-422. doi:10.1080/00131911.2018.1424116

Harrison, A., Smith, H., Botelho, A., Ottmar, E., & Arroyo, I. (June, 2020). *For good measure: Identifying student measurement estimation strategies through actions, language, and gesture*. In the International Conference of the Learning Sciences (ICLS) 2020 Proceedings, pp. 869-870. https://par.nsf.gov/servlets/purl/10167245

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.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579–588. https://doi.org/10.1037//0022-0663.93.3.579

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*(1), 114

Miller-Cotto, D., & Auxter, A. E. (2019). Testing the ecological validity of faded worked examples in algebra. *Educational Psychology*, 1-15.

Ottmar, E., & Landy, D. (2017). Concreteness fading of algebraic instruction: Effects on learning. *Journal of the Learning Sciences*, *26*(1), 51-78.

Scheiter, K. (2020). Embracing complexity in research on learning from examples and from problem solving. *Applied Cognitive Psychology*, *34*(4), 906-911.

Schwaighofer, M., Bühner, M., & Fischer, F. (2016). Executive functions as moderators of the worked example effect: When shifting is more important than working memory capacity. *Journal of Educational Psychology, 108*, 982–1000. https://doi.org/10.1037/edu0000115

Smith, H., Closser, A. H., Ottmar, E. R., & Chan, J. Y. C. (2022). The impact of algebra worked example presentations on student learning. *Applied Cognitive Psychology,* 1-15. https://doi.org/10.1002/acp.3925

Stokes, D. E. (2011). *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press.

Tempelaar, D. T., Rienties, B., & Nguyen, Q. (2020). Individual differences in the preference for worked examples: Lessons from an application of dispositional learning analytics. *Applied Cognitive Psychology*, *34*(4), 890-905.