Title

From the Lab to the Classroom: Effects of Embodied Pedagogies on Students' Learning of Statistical Concepts

Author

Zhang, Icy(Yunyi)

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

From the Lab to the Classroom:

Effects of Embodied Pedagogies on Students’ Learning of Statistical Concepts

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy in

Psychology

by

Icy Zhang

2024

ABSTRACT OF THE DISSERTATION

From the Lab to the Classroom:

Effects of Embodied Pedagogies on Students' Learning of Statistical Concepts

by

Icy Zhang

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2024

Professor James W. Stigler, Chair

The idea that people learn from sensorimotor experiences, whether through performing actions themselves or observing others, has garnered increasing attention from researchers in psychology, cognitive science, computer science, and education. In teaching and learning research, a key question is whether these sensorimotor experiences can help students acquire abstract concepts in complex domains. Past research has revealed promising evidence in various

domains such as mathematics and physics regarding the benefit of incorporating some sort of bodily actions into learning. However, our understanding of how different types of bodily experiences impact learning is still nascent. Questions remain about the effect, mechanism, and practical application of using embodied experiences to help learners learn abstract knowledge in complex domains. These inquiries lead to a series of laboratory experiments and classroom interventions that I will present across three chapters, each written as a discrete empirical article that either has been published or is in preparation for publication. Across three chapters, the work was conducted in the field of statistics and data science education, which was picked because the concepts are intrinsically abstract and difficult, but they simultaneously do not require a sophisticated mathematical background. The first question focuses on the effect of observing bodily actions. Whereas abundant evidence has demonstrated the effect of performing actions, the concept of observing actions is less explored. Would simply observing hands-on representations lead to an increase in learning? This question is answered in Chapter 1, a published work that demonstrates the efficacy of observing hands-on representations in improving students' understanding of randomness and the shuffle() function in R programming used to simulate randomness. The second question focuses on the mechanism underlying the effect—an embodied representation has more sensorimotor engagement and visuospatial concreteness than an abstract representation, but does sensorimotor engagement offer a unique benefit beyond visuospatial concreteness? Chapter 2 is a manuscript under review that reports on a laboratory experiment designed to isolate the effect of sensorimotor engagement. The findings suggest that sensorimotor engagement offers a unique benefit beyond visuospatial concreteness by helping learners develop more robust visuospatial representations. The last question relates to the practical application of different types of embodied interventions when we have learners with diverse levels of prior knowledge in the classroom. Theories in embodied cognition, along with other empirical evidence in both motor and learning domains, suggest that humans rely on their knowledge of their own bodies to understand other people's movements. This insight prompted me to ask whether learners' prior knowledge would moderate the type of embodied intervention (i.e. performing versus observing) on learning. The third Chapter reports on the design of a curriculum-linked embodied intervention to implement embodied activities over the entire school term of a college-level introductory statistics course. Students were randomly assigned the role of a performer or an observer. The findings provided support for the Perform-First hypothesis, showing that compared to observing, performing hands-on activities diminished the correlation between prior knowledge and post-test performance. Overall, this body of work extends the theory of embodied learning and offers practical insights for teachers and curriculum developers about how to implement embodied interventions into their educational materials and instructions.

The dissertation of Icy Zhang is approved.

Craig K. Enders

Keith Holyoak

Catherine M. Sandhofer

Ji Son

James W. Stigler, Committee Chair

University of California, Los Angeles

2024

Dedication

To my family, who first taught me the meaning of love.

To my friends and partner, who have surrounded me with it.

To my mentors, to whom I owe the person I have become.

To all devoted educators, alongside whom I firmly stand to fulfill the mandate of education and make this world a better place for the young and vibrant.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am profoundly grateful to have been guided by numerous remarkable educators, advisors, and collaborators throughout my academic journey. My family were my first teachers. My grandparents first introduced me to psychology and are forever my role models for being a dedicated and rigorous scholar. My parents respect and support all my decisions ever since I was a kid, giving me the chance to study in a country they have never gone to. I also want to thank my middle school English teacher, Wenlin Zhao, and high school history teacher, Jinhai Li, for showing me the breadth and depth of the world and encouraging me to set forth on a journey that led me to where I am today.

All of my accomplishments would not have been possible without the support of my advisor, Dr. Jim Stigler. His mentorship began the first day I walked into his office as a second-year undergrad. Till today, I still do not know why he would take the time to patiently teach me everything about research but I owe him a debt of gratitude and have since aspired to be a scholar and mentor like him to pay it forward. Thank you, Dr. Stigler, for inspiring me with your passion, wisdom and kindness. I also want to thank all my mentors, collaborators and colleagues in graduate school: Dr. Ji Son, Dr. Karen Givvin, Dr. Catherine Sandhofer, Dr. Keith Holyaok, Dr. Hongjing Lu, Dr. Carig Enders and Dr. Erica Cartmill, Dr. Laura Fries, Dr. Mary Tucker, Dr. Maureen Gray, Alice Xu, Erjing Zhang, Shuhao Fu. Each of you has been a beacon of inspiration and support. Lastly, I want to thank myself, for genuinely and fully enjoying the past four years of this journey and for still holding the same passion and dedication that drove me at the start of this journey.

Below, I want to acknowledge the contributions of each person in the three Chapters.

Chapter 1 is a version of a published article.

Zhang, I. Y., Tucker, M. C., & Stigler, J. W. (2022). Watching a hands-on activity improves students' understanding of randomness. *Computers & Education*, *186*, 104545.

All authors contributed to the conception and design of the project. I.Z. and M.C.T. prepared the study materials. I.Z. analyzed and interpreted the data, and drafted the manuscript. All authors provided input and revised the manuscript.

All authors approved the manuscript for submission.

Chapter 2 is a version of a manuscript under review.

Zhang, I. Y., Son, J. Y., & Stigler, J. W. (under review). Watching Hands Move Enhances Learning from Concrete and Dynamic Visualizations. *Journal of Experimental Psychology: General*.

I.Z. lead the conception of the project. All authors contributed to the design of the project. I.Z. collected, analyzed and interpreted the data. I.Z. drafted the manuscript. All authors provided input and revised the manuscript.

All authors approved the manuscript for submission.

Chapter 3 is a version of a manuscript in preparation for publication.

Zhang, I. Y., Son, J. Y., & Stigler, J. W. (in preparation). The Role of Prior Knowledge in Effects of Embodied Pedagogies on Learning.

I.Z. lead the conception of the project. All authors contributed to the design of the project. All authors contributed to the development of study materials. I.Z. collected, analyzed and interpreted the data. I.Z. drafted the manuscript. All authors provided input and revised the manuscript.

All authors approved the manuscript for submission.

Vita

EDUCATION

University of California, Los Angeles

Psychology, Ph.D. (Expected June 2024)

Psychology, B.A., and Economics, B.A. (2019)

PUBLICATIONS & PEER REVIEWS

Zhang, I., Izad, T., & Cartmill, E. (2024). Embodying Similarity and Difference: the Effect of Listing and Contrasting Gestures during U.S. Political Speech. *Cognitive Science* [IF: 2.92, Citation: 0]

Zhang, I., Son, J. Y., Blank, I. A., Stigler, J. W. (2024). Utilizing Dynamic and Embodied Visualization to Facilitate Understanding of Normal Probability Distributions. *Memory & Cognition* [IF: 2.5, Citation: 0]

Zhang, I., Gray, M., Cheng, X., Son, J., Stigler, J. W. (2023). Using Analogical Learning Strategies to Improve Learning from an Online Statistics Textbook. *Journal of Experimental Psychology, Applied.* [IF: 2.81, Citation:0]

Zhang, I., Tucker, M. C., & Stigler, J. W. (2022). Watching a hands-on activity improves students' understanding of randomness. *Computers & Education*, 104545. [IF: 15.58, Citation: 8]

Zhang, I., Givvin, K. B., Sipple, J. M., Son, J. Y., & Stigler, J. W. (2021). Hand placement affects students' learning of an abstract concept from video. *Cognitive Science*, *45*(2), e12940. [IF: 2.92, Citation: 24]

Xiong, C., Lee-Robbins, E., **Zhang, I.,** Gaba, A., & Franconeri, S. (2022). Reasoning Affordances with Tables and Bar Charts. *IEEE Transactions on Visualization and Computer Graphics*. [IF: 5.23, Citation: 5]

Eghterafi, W., Tucker, M. C., **Zhang, I**., & Son, J. Y. (2022). Effect of Feedback with Video-Based Peer Modeling on Learning and Self-Efficacy. *Online Learning*, *26*(2), 1-21. [IF: 1.50, Citation: 2]

Zhang, I., Son, J. Y., Stigler, J. W. (under review). Does Sensorimotor Engagement Offer Benefits Beyond Concreteness? *Journal of Experimental Psychology: General*

Zhang, I., Cheng, X., Jia, Y., Son, J. Y., & Stigler, J. W. (under review). Empowering Novice Programmers: Redesigning Documentation for Effective Informal Learning. *Journal of Computer Assisted Learning*

FELLOWSHIPS, AWARDS AND ACHIEVEMENTS

UCLA Dissertation Year Fellowship	2023-2024
UCLA Charles F. Scott Fellowship	2023-2024
UCLA Charles E. & Sue K. Young Graduate Student Fellowship	2023
Liu Yunghuo Bei Qui Memorial Fellowship	2023
UCLA Charles F. Scott Fellowship	2022-2023

Teaching Experience

Nine quarters of teaching experience at UCLA

Teaching Fellow	2023-2024
Teaching Associate	2021-2023
Teaching Assistant	2020-2021

Chapter 1

Watching a Hands-On Activity Improves Students'

Understanding of Randomness

Icy Zhang, Mary C. Tucker, James W. Stigler

University of California, Los Angeles

Abstract

Introductory statistics students struggle to understand randomness as a data generating process, and especially its application to the practice of data analysis. Although modern computational techniques for data analysis such as simulation, randomization, and bootstrapping have the potential to make the idea of randomness more concrete, representing such random processes with R code is not as easy for students to understand as is something like a coin-flip, which is both concrete and embodied. In this study, in the context of multimedia learning, we designed and tested the efficacy of an instructional sequence that preceded computational simulations with embodied demonstrations. We investigated the role that embodied hands-on movement might play in facilitating students' understanding of the shuffle function in R. Our findings showed that students who watched a video of hands shuffling data written on pieces of paper learned more from a subsequent live-coding demonstration of randomization using R than did students only introduced to the concept using R. Although others have found an advantage of students themselves engaging in hands-on activities, this study showed that merely watching someone else engage can benefit learning. Implications for online and remote instruction are discussed.

Keywords: hands-on demonstration, computer simulation, statistics education, multimedia learning, online instruction, instructional sequence, embodied cognition

Watching a Hands-On Activity Improves Students' Understanding of the Shuffle Function in R

A long-term challenge for statistics educators has been finding effective ways to help students understand randomness as a data generating process (Garfield & Ben-Zvi, 2005; Zieffler et al., 2008). Although hands-on activities such as coin flipping and dice rolling have long been considered an important part of the statistics educators' toolbox (Dyck & Gee, 1998; Lunsford et al., 2006), connecting such activities to important statistical concepts such as sampling distributions and hypothesis testing has proven difficult in practice (Pfaff & Weinberg, 2009). Students find it difficult to make these connections, which requires seeing a distribution of data as just one of many possible distributions that could have been produced by a random process.

Recent developments in the field of statistics and data science, however, provide new opportunities for students to apply concepts of randomness to the interpretation of data (Chance, & Rossman, 2006). Once almost entirely based on mathematics and mathematical models, statistics is increasingly becoming a computational science. Techniques such as simulation, randomization, and bootstrapping provide a less algebraic and thus relatively more concrete basis for understanding how simulations of randomness can be applied in the practice of data analysis (Pfaff & Weinberg, 2009). Instead of proving what a distribution of a sample statistic would look like under certain conditions using calculus, empirically simulating distributions of statistics under certain conditions and directly observing what they end up looking like is now possible with computer code.

A common but very difficult task in statistics is to imagine a circumstance, such as when two variables have only a random relationship, and then predict the resulting probability distributions of possible sample statistics (e.g., the correlation of the two variables).

Computational techniques such as randomization allow us to program up a simulation where two variables are randomly related, generate many samples, calculate the sample statistic of interest, and then examine the result of these simulated statistics. Not only do computer-based simulations support new methods of statistical analysis (e.g., randomization or permutation tests), but they also provide new ways of teaching students about randomness.

The shuffle function in R, part of the mosaic package (Pruim et al., 2017), allows students with minimal experience in coding to quickly and easily construct a randomized sampling distribution based on many random shuffles of an actual data set. Using the shuffle function, students can construct a sampling distribution of the difference between two experimental groups by repeatedly randomizing the pairings of grouping and outcome variables in a data set. The resulting sampling distribution of differences would be centered at 0, because any relation between group and outcome would be broken by the shuffling. The standard error of the distribution would give some indication of how likely various differences would be if the null hypothesis were true. And the sample statistic of interest—the observed difference between groups in an actual study—could be interpreted in the light of this sampling distribution.

A number of investigators have explored the use of such computational simulations to support students' understanding of statistical concepts (Chance & Rossman, 2006; Hodgson & Burke, 2000; Wood, 2005). Compared to traditional in-class simulation activities, such as coin-flipping, computer simulation offers several advantages. For example, computer simulations can be repeated very quickly, thus enabling students to see the results of many random iterations in a more concrete way than ever before (Hancock, & Rummerfield, 2020). Further, the results of simulations can be instantly represented in multiple modalities, such as tables and graphs, potentially resulting in more flexible understanding of complex concepts such as randomness

(Ainsworth & VanLabeke, 2004; Chance, & Rossman, 2006; Zhang, & Maas, 2019). And because computer simulations require only a computer and relatively little setup, they are more feasible to implement in large undergraduate classes than are traditional hands-on experiments, which require rolling pennies or dice over many iterations.

However, despite the potential of computer-based simulation methods, a review of the relevant literature suggests mixed evidence overall for the effectiveness of simulation as an instructional tool (Chance, et al., 2004; delMas et al., 1999; Lane, 2015). Though some studies show the benefits of simulation, others have shown that the use of computer simulations provides only limited benefit to students and can, in some cases, impede learning by exacerbating students' misunderstandings or increasing their level of confusion (Watkins et al., 2014). Other researchers have noted that despite statistically significant research findings of the effect of computer simulations, the observed increase in students' understanding was not substantial (e.g., delMas et al., 1999).

Computer simulations can provide experts with a fast and efficient way to explore various statistical scenarios. However, because such simulations are highly complex perceptual objects, they can be confusing for novices who do not know what they are looking at (e.g., is this a sample or a sampling distribution?) nor where to look during a dynamic simulation. Thus, simulations may potentially overload novices' working memory (Savinainen et al., 2005).

Working memory is a short-term system into which information from the environment flows before it is encoded into long-term memory (Baddeley, 1992). Because teaching randomness with computer simulations requires keeping track of multiple elements, students may have difficulty connecting particular components of a simulation with the new and abstract statistical concepts they are intended to learn. As a result, this kind of instructional experience imposes high demand on the learners' working memory (Sweller, 2010, 2020; Sweller et al., 2019) and depletes attentional resources (Tarmizi & Sweller, 1988).

While computer simulations can provide powerful demonstrations of key statistical concepts, students' attention may need to be directed and scaffolded in order for such simulations to be effective. In contrast, embodied and concrete activities, such as coin-flipping, are easier to understand and connect to learners' prior learning, but limited in their potential to quickly show patterns that can only be seen over thousands of iterations. An instructional sequence that combines the benefits of a more embodied approach with the benefits of computer simulations might help students connect simulations to their prior experience and to important statistical concepts.

The main goal of the work reported here is to design and test an instructional sequence that is solidly grounded in theories and findings from cognitive psychology, including work on cognitive load, embodied cognition, and the design of instructional sequences. We are especially interested in a body of research and theory known as "concreteness fading" (Fyfe & Nathan, 2019). According to this work, an instructional sequence in which concrete representations are introduced *before* abstract representations may maximize learning. This suggests that rather than choose between hands-on demonstrations and more abstract computer simulations, it might be best to do both, with the hands-on activity preceding the computer simulation.

According to the concreteness fading hypothesis, concrete representations more easily connect to prior knowledge, and then provide a foundation on which to build new, related abstract representations (Fyfe & Nathan, 2019; Glenberg et al., 2004; Goldstone & Son, 2005; Kokkonen & Schalk, 2021). For example, seeing physical pieces of paper being "shuffled" helps connect to students' prior experience of shuffling in the physical world (e.g., with cards), which might subsequently help with their understanding of a computational simulation that “shuffles” rows of a data frame.

By connecting the more abstract computer simulation with their everyday experience of shuffling, students' attention is constrained and directed toward the most relevant aspects of the computer simulation. Although some concreteness fading theories have proposed three progressive forms (i.e., enactive, iconic, and symbolic; e.g., Fyfe et al., 2014), in the current study, we focus simply on preceding a relatively less concrete experience with one that is more concrete. (Other studies of concreteness fading have followed a similar approach, e.g., Goldstone & Son, 2005).

Beyond the instructional sequence suggested by the concreteness fading hypothesis, we also connect our work to the broader literature on embodied cognition. This literature has clearly established that bodily movement can lessen cognitive load and support learning (Ballard et al. 1997; Paas & van Merriënboer, 2020; Pouw et al., 2014; Varga & Heck, 2017). For example, research has shown that both observing and performing gestures can provide a way to introduce and coordinate multiple pieces of information without increasing cognitive load, which in turn can benefit learning (Cook et al., 2013; Goldin-Meadow & Alibali, 2013; Goldin-Meadow et al., 2001; Rueckert et al., 2017). Gestures are beneficial not only because they temporarily offload information to the hands and physical space (Chu et al., 2014) but also because they provide another modality for representing information (Sepp et al., 2019). The modality effect in cognitive load theory states that simultaneously presenting information in more than one modality, such as adding in an embodied modality, increases working memory capacity beyond that available to one modality alone, thus expanding the cognitive resources available for learning (Paas & van Merriënboer, 2020).

Interestingly, the embodied cognition literature suggests that physical movements can shape cognition and learning even when students merely observe these movements (Da Rold, 2018; Tran et al., 2017). Neurons with mirroring properties have been shown to be activated both when performing and when watching others perform a similar physical action (Fu & Franz, 2014). More importantly, this mirroring only occurs when observing embodied human actions, not when observing disembodied ones such as ball movements. This suggests that an instructional sequence that leads with a more concrete and embodied experience may not require a physical hands-on activity. Benefits may occur from simply watching a hands-on demonstration.

The Current Study

The research reported here lies at the intersection of these research literatures: statistics education, cognitive load theory, the design of instructional sequences, and embodied cognition. Most relevant to the current study is a recent one by Hancock and Rummerfield (2020), in which students engaged in concrete, hands-on activities before engaging in computer-based simulations. The authors found a small yet significant effect in which students learned more about the concept of sampling distributions when instruction with simulation applets was preceded by a hands-on activity. However, in that study, students physically performed the hands-on activity themselves. Left unanswered was whether simply observing hands-on activities in a multimedia learning context could produce a similar effect.

In the current study, prompted by the shift to remote instruction during the COVID-19 pandemic, we investigated the same instructional sequence as Hancock and Rummerfield (2020), preceding computer simulation with a hands-on activity. But this time, instead of having students participate in a hands-on activity, we had them observe someone else engaging in the activity. It is hard enough to implement hands-on activities in large classes, but the prospect of doing so online seemed even more daunting. It would be of great practical significance if merely watching a video of a hands-on activity could enhance learning. Based on the argument postulated by the modality effect and the literature on embodied cognition, watching a hands-on demonstration could show a benefit similar to that found by observing gestures.

In this initial investigation, we randomly assigned participants into one of two groups: a *hands-on* group and a *live-coding* group. In the live-coding group, students watched a video of R code being typed and run on a screen as a narrator explained the workings of the shuffle function in R (Pruim et al., 2017). In the hands-on group, students watched a video with the same narration, but instead of watching someone code in R, they watched a pair of hands simulate the shuffle function by cutting and rearranging pieces of paper with data written on them. Both groups of students then watched the same live-coding video in which the shuffle function was used to create a sampling distribution. The verbal modality and visual modality were employed in both conditions, whereas the embodied modality was only present for the first video in the hands-on condition.

The question of interest to us was whether watching a hands-on simulation of the shuffle function prior to instruction using computer simulation would result in a better, more flexible, and more transferable understanding of the shuffle function (e.g., its use for creating sampling distributions and the interpretation of the resulting sampling distributions) than would simply watching someone explain the function as they entered and ran code in R. We report two studies with college students taking an introductory statistics class in a public research institution. The second study is primarily a replication of the first.

Study 1

Method

Participants

Thirty-three undergraduate students from a large public research university participated in the study. All students had completed the same introductory statistics course in the psychology department, taught by two different instructors using the same course material, during the previous academic quarter. Students from this class were chosen because they had a common set of background experiences relevant to the study—All had been taught how to use the shuffle function in R and had used the function to think about whether randomness alone could have generated a sample distribution (i.e., without the effect of an independent variable).

The two statistics instructors from the prior term emailed their former students to invite them to participate in the study. Students were told that their participation would help the textbook authors to improve the book for future classmates. Those who chose to participate were given a five-dollar gift card after completing the study. The study design, as well as our method for recruiting and compensating participants, was reviewed and approved by the university's institutional review board for the protection of human subjects.

Design & Procedure

The study was conducted through Qualtrics (https://www.qualtrics.com). On clicking the survey link, students were randomly assigned into one of two conditions: *hands-on* (n = 18) or *live-coding* (n = 15). Both versions of the survey were structured in the same way. Students first rated their attitudes toward programming in R, then answered two free-response questions designed to assess what they remembered about the shuffle function in R from their course. Next, they watched a series of two videos about the shuffle function and the concept of randomness.

The first video contained the same instructional content across the two groups, but differed in the mode of presentation depending on which group students had been assigned to. The *hands-on* video showed an instructor's hands manipulating a dataset on paper, cutting and shuffling the pieces of data, much as might occur during an in-class hands-on exercise. The *live-coding* video showed a screen recording of an instructor writing and running R code in an interactive online Jupyter notebook.

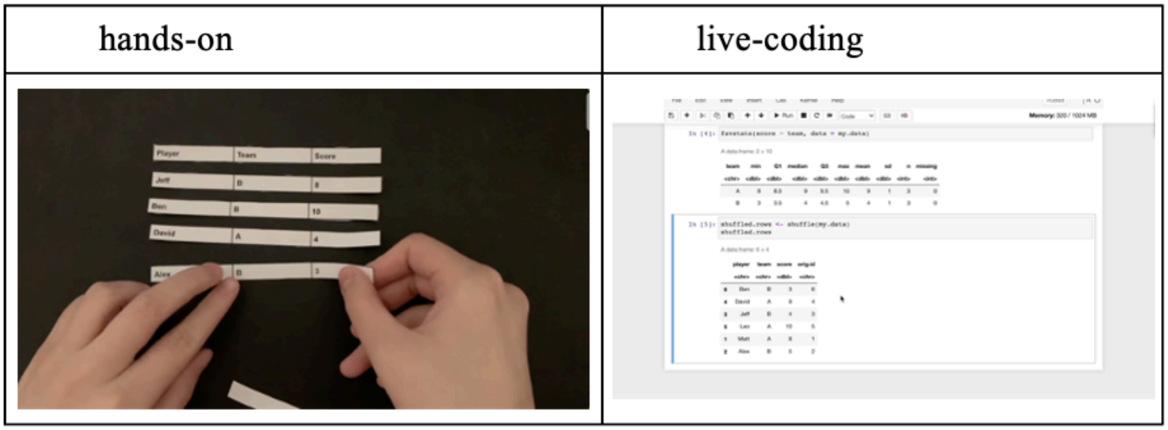
The second video was identical across the two conditions. Students watched an instructor write code in R and think aloud as they worked through a series of R commands (similar to the first video in the *live-coding* condition). After watching each video, participants rated how difficult it was to comprehend. At the end of both videos, participants completed a 22-question survey that assessed their understanding of the video and its contents and their perceptions of the activity (e.g., how much they liked the videos).

Materials

The two videos shown first (one hands-on, the other live-coding) were matched in content. Both videos explained how the shuffle function works. In the live-coding condition, participants watched a narrator type and run R code in a Jupyter notebook while explaining what they were doing out loud. (Figure 1). The narrator used the shuffle function to shuffle one variable in a small data set. In the hands-on condition, participants watched a person cut a printed data table into pieces and then rearrange those pieces randomly, simulating exactly what the shuffle function did in the live-coding video. As they manually shuffled the pieces of data, the narrator explained what they were doing, using almost identical language as used in the live-coding video.

Figure 1

Screen Grabs from the Hands-On Video and the Live-Coding Video



The only difference in narration across the two videos was in the language used to describe shuffling. For example, in the hands-on condition, the instructor would shuffle the data by physically moving the pieces of paper and say, “We can see as we shuffle the rows, the position of each row changes. For example, Matt started in position 1, but moved to a different position after we shuffled.” In the live-coding condition, the instructor would write down the R code, then press run and say, “When we shuffle the rows, R creates a new variable called orig.id. This tells us what position each row occupied in our original dataset. For example, Matt has an orig.id of 1. This is because Matt was in row 1 of our original dataset.” Then, in both conditions, the instructor would ask rhetorically, “Is that what you expected it would do? Why or why not?”

The hands-on version of the video was recorded by placing a camera so as to look down from above at the hand movements of the instructor. The live-coding video was created via a screen recording of the instructor typing and running code in a Jupyter notebook (Kluyver et al., 2016). The second live-coding video (common across the two conditions) was similar in format to the first live-coding video.

The second video, identical across conditions, was a live-coding video that involved applying concepts learned in the first video to a larger dataset adapted from a real experiment. The dataset (called the laptop dataset) involved one independent variable (whether students viewed a laptop screen during class) and two dependent variables (students' final grades and students' self-rated level of distraction). In the video, the instructor used the shuffle function in R to explore whether there was an effect of condition on these two outcome measures.

Measures

Pre-survey and pretest. The pre-survey measures asked students how they felt about their R skills, whether they learned shuffle in their class and asked them to rate, on a scale of 0 to 10, how well they understood the shuffle function. The pretest contained two open response questions: "In your own words, explain what the shuffle() function does." and "In your own words, explain when you would use the shuffle() function." The purpose of the pretest was to make sure, given the small sample size, that the two experimental groups did not differ in their understanding of the shuffle function prior to watching the videos.

Posttest and post-survey. The posttest contained 22 questions designed to assess students' understanding of the shuffle function and the concept of randomness. It also included transfer questions that asked students to make and interpret statistical inferences. For example, in one question, students were shown one shuffled and one non-shuffled faceted histogram and asked to reason about whether there could be a difference between the two conditions. It asked again at the end of the test, "What do you think the purpose of the shuffle() function is?" and "In your own words, explain when you would use the shuffle() function."

Each correct response was awarded a maximum of one point, with possible scores ranging from 0 to 22. A partial credit of 0.5 was given to answers that were partially correct but were missing pieces or manifested some minor misunderstandings. The scoring of the free-response questions were conducted by two trained research assistants. They coded the questions based on a predetermined rubric, blind to condition. For each question, the discrepancy rate between the two research assistants was lower than 10%. Then, the two research assistants met to discuss the discrepancies until a consensus was reached.

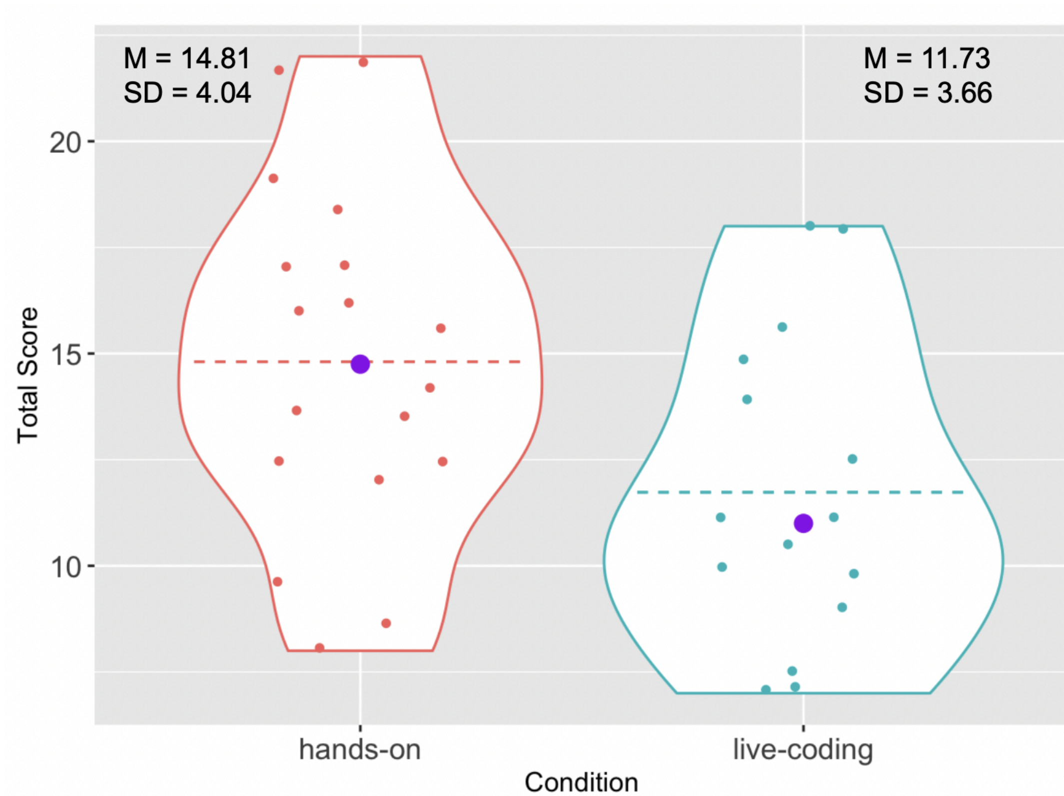
In the post-survey, students again were asked to rate, on a scale of 0 to 10, how well they understood the shuffle function. A change in self-rated understanding score was computed by subtracting the pretest rating of understanding from the posttest rating. Students also were asked, using a Likert scale (from strongly disagree to strongly agree), how much they agreed with statements expressing that "they would like to see more activities like this in their own online textbook," "they liked this way of learning R," and "they learned a lot from the activity."

Results

An analysis of pretest scores found no significant difference across conditions in students' prior understanding of the shuffle function ($t(31) = .17$, $\eta^2 = .00$, 90% CI = [.00, .06], $p = .864$).

Figure 2

Violin Plot Showing Posttest Scores by Condition



Note. Dashed lines show the mean of each group. Purple dots show the median.

Figure 2 shows overall posttest scores by condition. Participants in the hands-on condition performed better on average on the posttest than participants in the control condition ($t(31) = 2.27$, $\eta^2 = .14$, 90% CI = [.01, .34], $p = .031$). Similar benefits of the hands-on group were observed when pretest was included as a covariate in the multiple linear regression model ($t(30) = 2.23$, $\eta_p^2 = .15$, 90% CI = [.01, .34], $p = .033$). When included as a covariate (i.e., controlling for condition), students' pretest performance did not predict posttest scores ($t(30) = .86$, $\eta_p^2 = .02$, 90% CI = [.00, .21], $p = .396$).

Independent t tests for each question revealed two open response questions on which the hands-on group performed better than the control group. These questions asked students to 1) explain what would happen to the number of observations in one condition if the condition variable were shuffled ($t(31) = 2.35$, $\eta^2 = .15$, 90% CI = [.00, .38], $p = .025$); 2) describe what a specific line of code that shuffled the outcome variable in the dataset was doing ($t(31) = 2.06$, $\eta^2 = .12$, 90% CI = [.01, .35], $p = .048$). One free response question that asked students to imagine and describe how a histogram would be different if one of the variables were shuffled prior to running the code yielded some group difference but the difference in this question was not statistically significant ($t(31) = 1.96$, $\eta^2 = .11$, 90% CI = [.00, .34], $p = .059$).

Next, we examined whether participants' self-rated understanding of the shuffle function before and after watching the videos differed by condition. The difference between conditions was not statistically significant ($t(31) = 1.30$, $\eta^2 = .05$, 90% CI = [.00, .23], $p = .204$). We also examined if participants would like to see more activities like this in their textbook. A linear regression showed that the difference between the two conditions was not statistically significant ($t(26) = .40$, $\eta^2 = .00$, 90% CI = [.00, .13], $p = .691$).

To evaluate the impact of the intervention on students' metacognition, we explored the relationship between students' self-rated understanding of the shuffle function post intervention and their performance on the posttest. A linear regression showed that students' self-rated understanding post intervention was a significant predictor of their posttest performance ($t(31) = 2.05$, $\eta^2 = .12$, 90% CI = [.00, .35], $p = .049$). However, students' change in self-rated understanding from pre to post intervention did not significantly correlate with performance ($t(31) = 1.29$, $\eta^2 = .05$, 90% CI = [.00, .26], $p = .207$).

Discussion

In this study, we found preliminary evidence that preceding a live-coding video with one showing a hands-on simulation of the shuffle function can improve students' understanding of the shuffle function and the concept of randomness. The study is, to our knowledge, the first to test experimentally if students benefit from embodied experiential learning in a concrete to abstract instructional sequence when their participation is limited to watching a video of someone else engaging in a hands-on experience. It is important to note that students' participation was completely online in both the hands-on and live-coding conditions; in both groups, students' participation only involved watching instructional videos.

Because we used a live-coding video as the control, the findings suggest that it is something specific about seeing the hands carry out the randomization, not just the "in the moment" nature of the demonstration, that benefits learning. Our result lines up with many studies in the gesture literature that have found that learning is enhanced even when learners were merely observers of gestures during learning (Cook et al., 2013; Rueckert et al., 2017; Son et al., 2018). For example, Cook et al. (2013) found that observing hand gestures during mathematical learning benefited students' immediate and delayed posttest performance.

The findings also make sense in relation to the theory of embodied cognition and the modality effect in cognitive load theory. Watching a video of hands shuffling pieces of paper offers an additional modality (i.e., the embodied spatial modality) to the multimedia learning context in addition to the visual and auditory modalities. This added modality may have activated embodied representations of the core ideas that underlie the shuffle function and eased the cognitive load by providing another pathway for students to take in and process information in addition to the already active pathway of language processing.

The efficacy of this instructional sequence with embodied activities and computer simulation casts light on the teaching of statistics and computer programming in the digital era. Practically, given the growing interest in using statistical programming languages like R as pedagogical tools, the findings of this study provide important and encouraging insights into the use of hands-on demonstrations to complement computer simulation in remote teaching.

This study shows promising evidence that students can benefit from embodied hands-on experiential learning even when they are just observers of the activity. Nevertheless, it is important to keep in mind that this study is still exploratory and is limited by its small sample size. We set out to replicate the findings from Study 1 with a larger sample of students in Study 2.

Study 2

Method

Participants

Based on the results of Study 1, we conducted a power analysis to determine the sample size needed for the replication. Given an η^2 effect size of around .14, obtaining a power of .7 or .8 required a sample size of 20 or 25 participants per group.

Forty-seven undergraduate students taking introductory psychological statistics during a summer session at the same public research institution participated in the study. Participants were between the ages of 18 and 23 (M = 19.89, SD = 1.09) and 53.19% identified as Asian, 8.51% Black or African, 25.53% White, 2.13% American Indian or Alaska Native American, and 23.40% other. Students were emailed a link to the survey and told they would receive extra credit toward their course grade if they completed the survey. Given the sample size, the power of this replication is between .7 to .8. As before, the study design and procedures were approved by the institutional review board for protection of human subjects.

Design, Procedure, and Measures

The design and procedures for Study 2 were identical to those used in Study 1. On clicking the survey link, students were randomly assigned into one of the two conditions: *hands-on* (n = 20) or *live-coding* (n = 27). Students answered the same pre-survey questions and posttest items and watched the same series of videos as in Study 1.

The posttest included all 22 questions used in Study 1, plus 9 additional open-ended questions designed to probe students' explanations for their multiple choice answers and to assess transfer beyond the content covered in the video. Each question was given a maximum of one point, with possible scores ranging from 0 to 31. The post-survey of attitudinal measures was identical to the one used for Study 1.

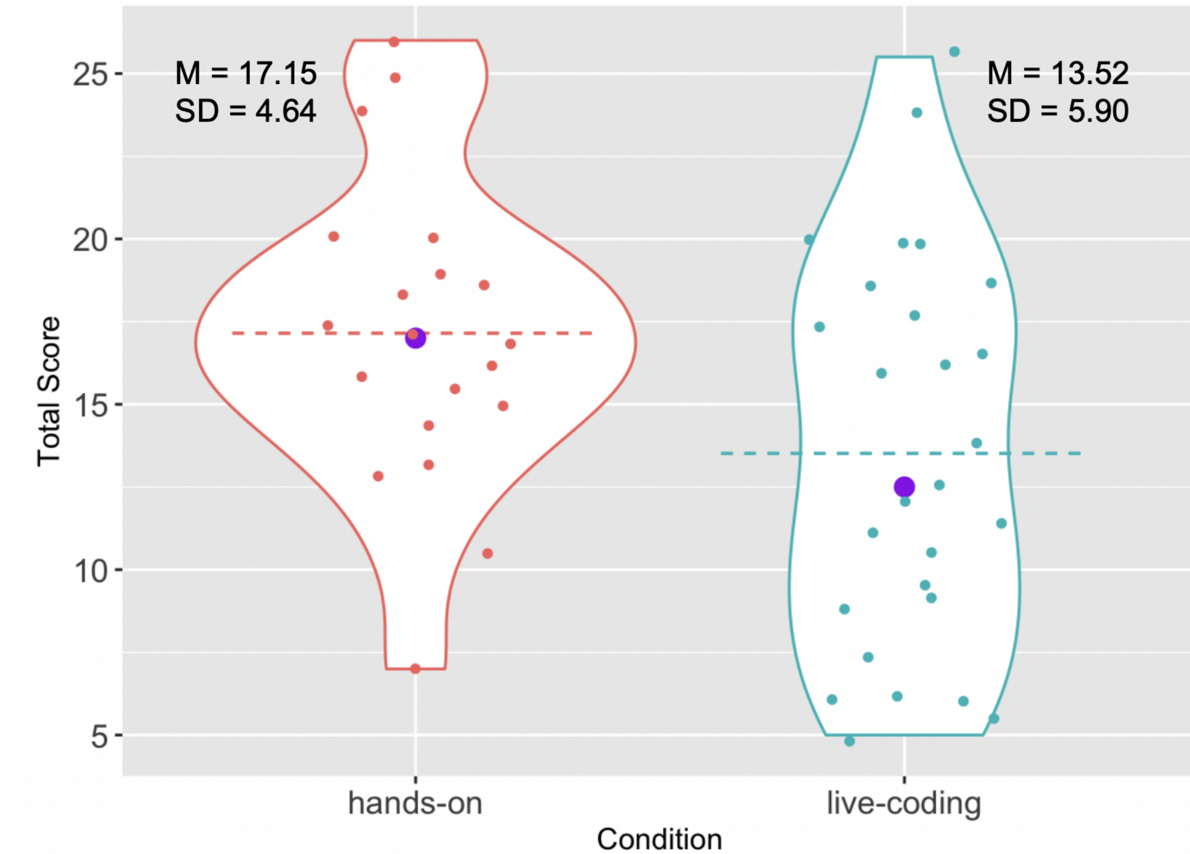
Results

We conducted a two-tailed independent t test to examine if there were any pre-existing differences between the two conditions. The two groups did not differ significantly from one another on the pretest ($t(45) = -.07$, $\eta^2 = .00$, 90% CI = [.00, .00], $p = .945$).

Figure 3 shows the distribution of participants' posttest scores by condition. Replicating the results of Study 1, participants in the hands-on group performed better on average than participants in the live-coding group ($t(45) = 2.28$, $\eta^2 = .10$, 90% CI = [.01, .26], $p = .028$). As in Study 1, this difference remained statistically significant when controlling for students' performance on the two-question pretest using a multiple linear regression ($t(44) = 2.35$, $\eta_p^2 = .11$, 90% CI = [.01, .27], $p = .023$).

Figure 3

Participants' Performance on Posttest by Condition



Note. Dashed lines show the mean of each group. Purple dots show the median.

Independent t tests for each question revealed two open response questions and one multiple-choice question, for which the hands-on group performed better than the control group. On the multiple-choice question, students were shown R code that shuffled the condition variable in a dataset and were asked what effect they thought running the code would have on the value of condition for row 1 of the data set ($t(45) = 3.80$, $\eta^2 = .24$, 90% CI = [.06, .44], $p < .001$). The free-response questions with significant group effects: 1) showed students the code to create a faceted histogram with an actual dataset and asked them whether the group difference visible in the graph could be due to randomness ($t(45) = 2.96$, $\eta^2 = .16$, 90% CI = [.02, .36], $p = .005$); 2) showed students the code to create a faceted histogram with shuffled data and asked them what might have caused the difference in the means represented in the graphs ($t(45) = 2.61$, $\eta^2 = .13$, 90% CI = [.01, .32], $p = .012$).

As in Study 1, participants' change in self-rated understanding of the shuffle function as a result of watching the videos did not differ across conditions ($t(31) = 1.39$, $\eta^2 = .04$, 90% CI = [.00, .18], $p = .173$), nor did their ratings of how much they would like to see more activities like this in the future ($t(40) = 1.26$, $\eta^2 = .04$, 90% CI = [.00, .21], $p = .216$). Also as in Study 1, linear regressions showed that students' post-intervention ratings of understanding significantly predicted performance on the posttest ($t(41) = 2.54$, $\eta^2 = .14$, 90% CI = [.00, .34], $p = .015$), whereas participants' change in self-rated understanding from pre to post intervention did not significantly predict posttest performance ($t(41) = 0.19$, $\eta^2 = .00$, 90% CI = [.00, .08], $p = .851$).

General Discussion

In both Study 1 and Study 2, students who watched a hands-on video before a live-coding video performed better on the posttest than students who watched two live-coding videos. Interestingly, despite learning more, students in the experimental group did not necessarily believe they learned more or enjoyed the experience more. Notably, the effect did not involve students themselves engaging in a hands-on activity, but only watching someone else engage in the activity on an instructional video. Together, these two studies demonstrate the efficacy of an instructional sequence in which computer simulation is preceded by embodied movements to support learning.

We think this instructional sequence that precedes computational simulation with hands-on demonstrations is beneficial for two reasons. First, it is possible that the hands-on video made the shuffle function and the concept of randomness more concrete. According to concreteness fading and cognitive load literature, the embodied representations help offload some cognitive processing to the embodied modality and help connect to learners' experience in the physical world, thus reducing cognitive load and improving learning (Weisberg, & Newcombe, 2017). Previously occupied cognitive resources are thus freed up to process more information and later engage in problem-solving and inferences-making (e.g., Kastens et al., 2008).

Although the previous literature in embodied cognition has often focused on learners physically performing the actions themselves, findings from the gesture literature, especially the idea that merely observing the actions could be beneficial as well, align with the results of our studies. For example, research has shown that learners who observed the instructor's co-speech gestures about mathematical concepts achieved superior learning outcomes (e.g., extracted more useful information) than learners who did not see those gestures (Alibali et al., 1997; Goldin-Meadow et al., 1992).

In addition, based on the modality effect from cognitive load theory, it is possible that simply having more ways of representing information, especially during tasks that already require split attention, increases learning. The multiple representations literature would also suggest that having multiple representations (hands-on + live-coding) is better than having one representation (live-coding alone). Previous studies have found that being exposed to multiple representations of the same concept benefits students' learning in STEM domains (Acevedo Nistal et al., 2009; Cheng, 1999), because, according to cognitive flexibility theory, having more than one representation helps learners achieve a more adaptive and flexible knowledge reconstruction, which is a crucial feature of deep and transferable understanding (Spiro, 1988).

The current studies suggest a closer connection between the cognitive architecture put forward by cognitive load theory, the embodied cognition literature, and the instructional sequence literature. Whereas the previous cognitive framework in cognitive load theory primarily focuses on gestures, these two studies suggest that the active ingredient that improves learning may not be limited to gestures, but also includes arm movement and object manipulation. Although previous interventions in the literature concerning bodily movements beyond gestures have produced mixed results or small effect sizes, our studies consistently demonstrated a medium-to-large effect size of watching a hands-on demonstration.

Another interesting point to consider is that, despite the experimental group learning more, students did not differ significantly across conditions in how much they liked the intervention and their change of self-rated understanding. This finding makes sense considering that students are not known to be good judges of their own learning. Students often make such judgments based on heuristics in the study phase (Koriat, 1997), and their judgments are often influenced by processing fluency (Kornell et al., 2011), which is their subjective experience of how much effort they expended on processing information during learning (Alter & Oppenheimer, 2009). It suggests that the benefits of this intervention may not be perceivable to students.

Given a larger sample size, it would be interesting to know whether students were accurate in their ratings—for example, for students who rated their understanding as having decreased from pre- to post-intervention, did they in fact, perform worse on the posttest than they did on the pretest, and is that true across conditions? In addition, given that judgments of learning can be affected by processing fluency, are there students whose self-rated understanding decreased but whose performance actually improved from pre- to post-intervention?

The study delivers a practical and timely message to teachers as they work to plan their post-COVID-19 instructional activities as well as to those seeking to design better instructional videos with better instructional sequences. It validates the importance of giving students some hands-on exposure to the simulation processes prior to the computational simulation we want them to understand and also makes it clear that at least some of the benefits of embodied activities can be retained even if students are not performing the hands-on activities themselves. For instructors who are limited by class sizes, COVID-19 restrictions, or even simply class time, this study points to another possibility to utilize hands-on activities in instruction.

We also want to highlight the significance of the practice of instruction used in the current study, regardless of conditions. Traditional approaches in teaching statistics often emphasize computation and procedures while putting less emphasis on the importance of statistical thinking (Garfield & Ben-Zvi, 2005). Although a focus on memorizing the procedural steps to perform different statistical routines is a common method of teaching statistics, it often does not lead to transferable understanding (Fries et al., 2021). The instructional videos used in the two studies engaged students with statistical thinking and inferences instead of pieces of procedures, which would limit our capacity to foster students' ability to think and reason flexibly with unfamiliar data in new contexts.

This study explored a new method for instructors to promote students' informal statistical inferences. Through a combination of hands-on simulation and computer simulation, students were able to better recognize the omnipresence of variability, understand randomness and uncertainty, and use statistical methods to model them. This approach makes computer simulation more understandable for students with lower coding knowledge and fosters one crucial topic in informal statistical reasoning: reasoning with uncertainty and randomness.

Students are known to view statistics as a branch of mathematics and thus expect instruction to focus on numbers, formulas, and procedural computations with one unique right answer (Garfield & Ben-Zvi, 2005). However, if students view statistics as a set of procedures to achieve the correct answer, they are likely to feel uncomfortable thinking about variation and uncertainty in data. They are also less likely to consider randomness as a possible explanation for observed differences or patterns, a key component of statistical inference. Giving students exposure to embodied demonstrations prior to computational simulations may help them better appreciate uncertainty and randomness by shifting their attention from the output or conclusions of statistical tests to the processes that generate the data.

Limitations and Future Directions

The two studies reported here offer significant practical implications, but also bear some limitations to be addressed by future studies. One important next step to further extend our theoretical understanding of the mechanism is to add in a condition with students' own physical manipulation of the objects, and compare it against the current two conditions. It will be informative to know whether students' own physical actions would further benefit their learning above and beyond the benefits of observing the hands-on demonstration due to increasing level of embodiment or physically manipulating the objects themselves will actually be too cognitively demanding (i.e., adding too much extraneous cognitive load) that their learning would fall behind the group who observed the physical manipulation only.

Another important condition to consider is a condition with the same object manipulation as the hands-on condition but without the actual hands. Future studies should examine this condition because that would help distinguish two competing explanations for the observed improvement of understanding in our studies, whether it is through an activation of an embodied pathway or through simply the concreteness in the object manipulations. If embodied cognition is truly the explanation, the condition without the actual hands would be inferior to the hands-on condition. Moreover, future studies should explore ways to measure students' level of embodiment after the intervention to examine if an elevation of the level of embodiment is truly the mechanism.

In summary, the two studies reported here leveraged findings from multiple literatures in cognitive psychology to design and test the efficacy of an embodied-to-abstract instructional sequence to improve students' understanding of randomness, their use of R functions to simulate randomness, and their subsequent statistical inferences. It bears an important practical message for statistics education and also directs future research to promising advances in our theoretical understanding of the field of embodied cognition, cognitive load, and instructional sequence.

Appendix A

Pre-Test Attitudinal Measures

1. In Psych 100A, you learned how to do some R programming. How are you feeling about your R skills?
   a. Extremely bad
   b. Neither good nor bad
   c. Somewhat bad
   d. Somewhat good
   e. Extremely good

2. Did you learn about the shuffle() function in R in your Psych 100A class?
   a. Yes
   b. No
   c. Not sure / can't remember

3. How well do you understand what the shuffle() function does? (from 0 to 10, with 0 being not at all)

Pre-Test Questions

1. In your own words, explain what the shuffle() function does.

2. In your own words, explain when would you use the shuffle() function.

Post-Test Questions

The laptop_data dataset contains data from an experiment on the effect of laptops on student learning. Undergraduate students were randomly assigned to one of two conditions: view or no-view. In the view condition, students attended a 40 minute lecture and were allowed to keep their laptops open. In the no-view condition, students attended the same lecture, but were asked to keep their laptops closed. At the end of the lecture, students took a test on the lecture content and rated how distracted they felt during class.

There are three variables in this dataset:

- condition: the condition students were randomly assigned to, either view or no-view
- total: the percentage of questions students answered correctly on the post-lesson assessment
- distracted: students’ self-reported rating of how distracted they were in class.

1. What would you expect to happen to the value of **condition** for row 1 if we ran the code below?

   ```
   laptop_data$condition <- shuffle(laptop_data$condition)
   ```

2. What would you expect to happen to the value of condition for row 1 if we instead ran the code below?

   ```
   laptop_data$total <- shuffle(laptop_data$total)
   ```

We ran this code to create a table that shows the number of observations in each condition.

```
tally(~ condition, data = laptop_data)
```

```
condition
no-view    view
     19      19
```

Now, imagine we run this code:

```
laptop_data$condition <- shuffle(laptop_data$condition)
tally(~ condition, data = laptop_data)
```

3. What would happen to the number of observations in the view condition?
   a. The number of observations would increase
   b. The number of observations would stay the same
   c. The number of observations would decrease
   d. The number of observations would increase, decrease, or stay the same, but it's impossible to tell which

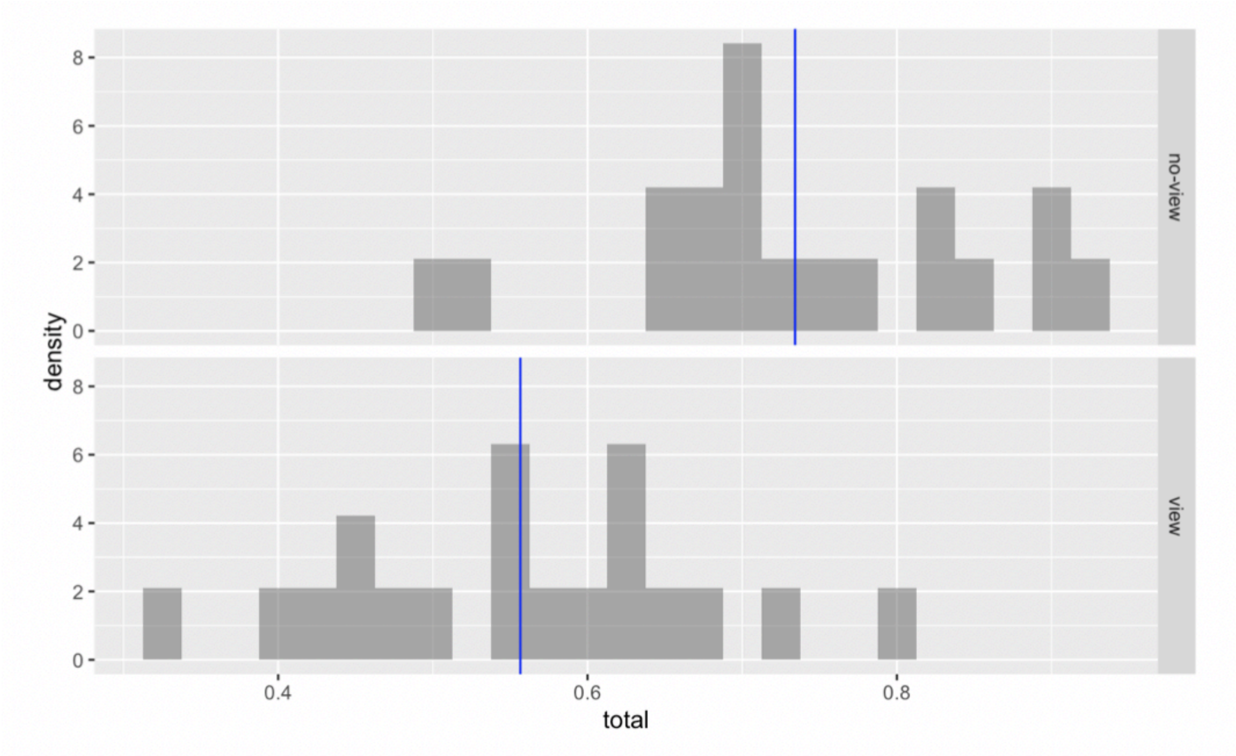
4. Explain your answer to the previous question

We used the code below to create a faceted histogram showing the distribution of total in each condition. The vertical lines represent mean total scores for the two conditions. Again, you can see that the participants in the no-view group scored higher, on average, than participants in the view group.

```
stats <- favstats(total ~ condition, data = laptop_data)

gf_dhistogram(~ total, data = laptop_data) %>%
```

```
gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
gf_facet_grid(condition ~ .)
```



5. Sometimes groups differ just because of randomness. Do you think the group difference in the histogram above could be due to randomness?
   a. Yes, it must be due to randomness
   b. No, it cannot be due to randomness
   c. Maybe, need to further investigate

6. If you wanted to investigate whether this difference could be due to randomness, what would you do?

   Please be as specific as possible in your response.

Take a look at each line of code below. For each line, **explain 1) what the code is doing** and 2) **why someone would write that code**.

```
laptop_data$condition.shuffle <- shuffle(laptop_data$condition)
```

7. What is this line of code doing?

8. Why would someone write this line of code?

```
laptop_data$total.shuffle<- shuffle(laptop_data$total)
```

9. What is this line of code doing?

10. Why would someone write this line of code?

11. Look at the two examples of codes below. Example 1 and Example 2 each produces a faceted histogram. In what ways would the two faceted histograms be similar? In what ways would the two faceted histograms be different?

Example 1:

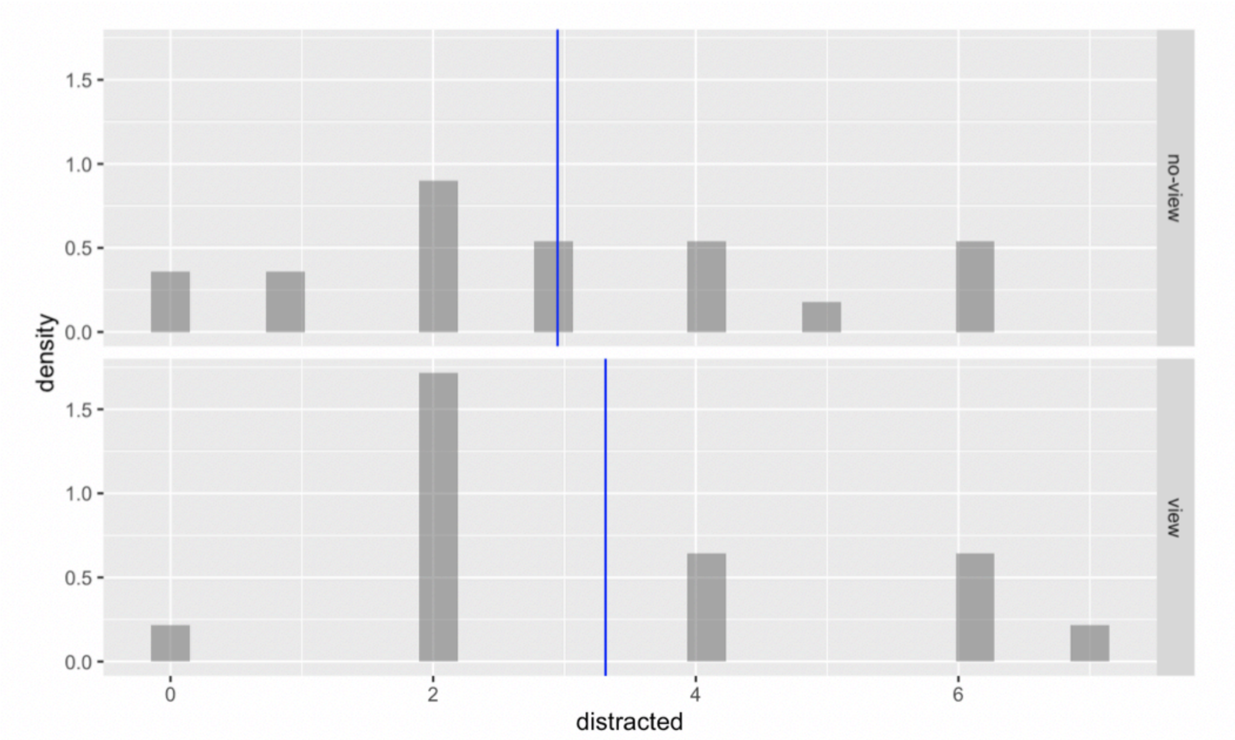
```
gf_dhistogram(~ distracted , data = laptop_data) %>%
  gf_facet_grid(shuffle(condition) ~ .)
```

Example 2:

```
gf_dhistogram(~ shuffle(distracted) , data = laptop_data) %>%
   gf_facet_grid(shuffle(condition) ~ .)
```

We ran this code to create the graph below. We added a line in each condition to represent the mean of **distracted** of that **condition**. Notice that the average **distracted** rating in the **no-view condition** is lower than the average **distracted** rating in the **view condition**.

```
stats <- favstats(distracted ~ condition, data = laptop_data)
gf_dhistogram(~ distracted, data = laptop_data) %>%
gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
gf_facet_grid(condition ~ .)
```



12. Sometimes groups differ just because of randomness. Do you think the group difference in the histogram above could be due to randomness?
    a. Yes, it must be due to randomness
    b. No, it cannot be due to randomness
    c. Maybe, we need to further investigate

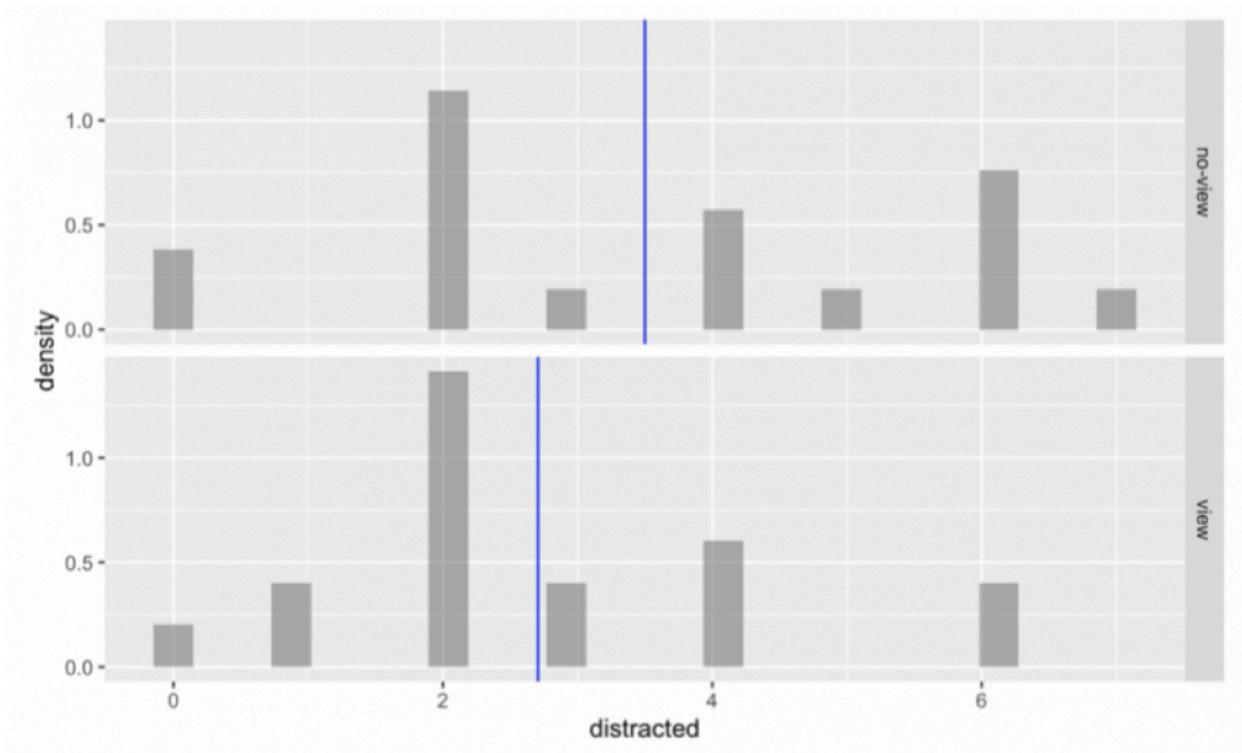
13. If you ran the code in the previous question again, do you think it would produce the same output?
    a. Yes
    b. No
    c. It's possible, but not likely

We revised the code from the previous question to create the graph below. We added a line to represent the mean of **distracted** for each **condition**. Notice that the average **distracted** rating in the **no-view condition** is higher than the average **distracted** rating in the **view condition**.

14. What caused the difference in the means represented in the graphs below?

```
laptop_data$condition.shuffle <- shuffle(laptop_data$condition)
stats <- favstats(distracted ~ condition.shuffle, data = laptop_data)
```

```
gf_dhistogram(~distracted, data = laptop_data) %>%
 gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
 gf_facet_grid(condition.shuffle ~ .)
```

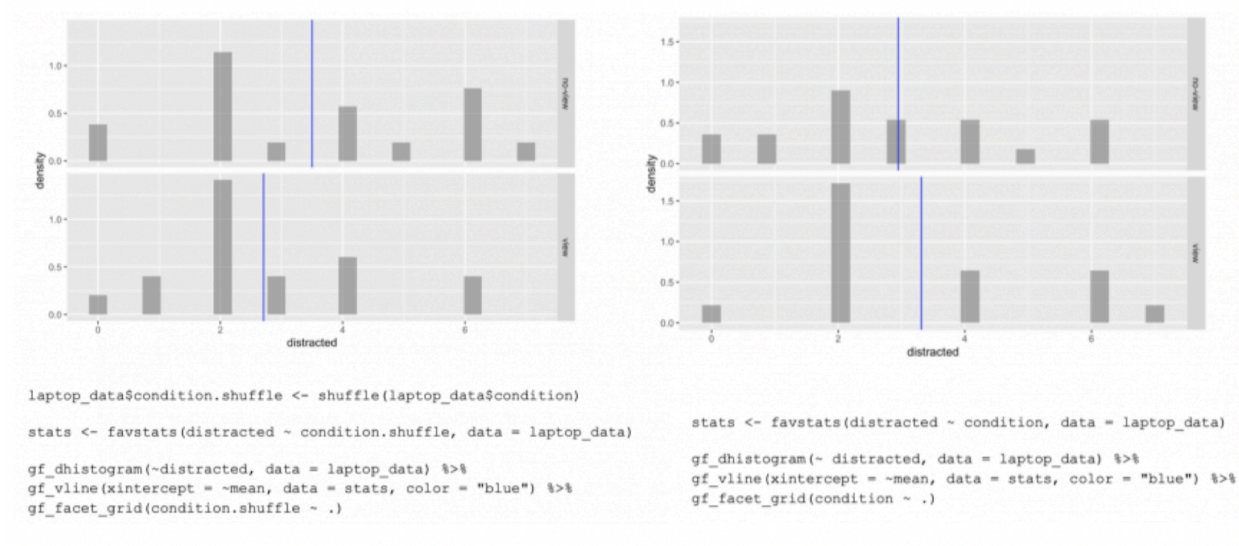


15. If you ran the code in the previous question again, do you think it would produce the same output?
    a. Yes
    b. No
    c. It's possible, but not likely

16. Explain your answer to the previous question

17. Sometimes groups differ just because of randomness. Do you think the group difference in the histogram above could be due to randomness?
    a. Yes, it must be due to randomness
    b. No, it cannot be due to randomness
    c. Maybe, need to further investigate

Look at the two faceted histograms below, along with the code that produced each:



18. Why do the two faceted histograms look different?

19. Based on what you've learned from these two histograms, do you think being able to view or not view a laptop during class (condition) affects students' self-reported rating of how distracted they were in class (as measured by distracted score on a post-lesson assessment)?

Imagine we run the code below:

```
laptop_data$distracted.shuffle <- shuffle(laptop_data$distracted)

mean(laptop_data$distracted.shuffle)
mean(laptop_data$distracted)
```

20. How would the mean of distracted.shuffle compare to the mean of distracted?
    a. The mean of distracted.shuffle would be larger
    b. The mean of distracted.shuffle would be smaller
    c. The two means would be the same
    d. It's impossible to tell

21. What do you think the purpose of the shuffle() function is?

22. In your own words, explain when would you use the shuffle() function.

Post-Test Attitudinal Measures

1. How well do you understand what the shuffle() function does? (from 0 to 10, with 0 being not at all)

Please rate your level of agreement with each of the following statements:

2. I learned a lot from this activity
   a. Strongly agree
   b. Agree
   c. Somewhat agree
   d. Neither agree nor disagree
   e. Somewhat disagree
   f. Disagree
   g. Strongly disagree

3. I like this way of learning R functions
   a. Strongly agree
   b. Agree
   c. Somewhat agree
   d. Neither agree nor disagree
   e. Somewhat disagree
   f. Disagree
   g. Strongly disagree

Appendix B

Pre-Test Attitudinal Measures

1. On a scale of 1 to 10, how math anxious are you?

2. In Psych 100A, you learned how to do some R programming. On a scale of 1 to 6 (with 1 being not at all confident and 6 being extremely confident), how confident do you feel in your R skills?

3. Did you learn about the shuffle() function in R in your Psych 100A class?
   a. Yes
   b. No
   c. Not sure / can't remember

4. On a scale of 1 to 10, how well do you understand what the shuffle() function does?

Pre-Test Questions

1. What do you think the purpose of the shuffle() function is?

2. In your own words, explain when would you use the shuffle() function.

Post-Test Questions

The laptop_data dataset contains data from an experiment on the effect of laptops on student learning. Undergraduate students were randomly assigned to one of two conditions: view or no-view. In the view condition, students attended a 40 minute lecture and were allowed to keep their laptops open. In the no-view condition, students attended the same lecture, but were asked to keep their laptops closed. At the end of the lecture, students took a test on the lecture content and rated how distracted they felt during class.

There are three variables in this dataset:

- condition: the condition students were randomly assigned to, either view or no-view
- total: the percentage of questions students answered correctly on the post-lesson assessment
- distracted: students' self-reported rating of how distracted they were in class.

1. What would you expect to happen to the value of **condition** for row 1 if we ran the code below?

```
laptop_data$condition <- shuffle(laptop_data$condition)
```

2. What would you expect to happen to the value of condition for row 1 if we instead ran the code below?

```
laptop_data$total <- shuffle(laptop_data$total)
```

We ran this code to create a table that shows the number of observations in each condition.

```
tally(~ condition, data = laptop_data)
```

```
condition
no-view    view
     19      19
```

Now, imagine we run this code:

```
laptop_data$condition <- shuffle(laptop_data$condition)
tally(~ condition, data = laptop_data)
```

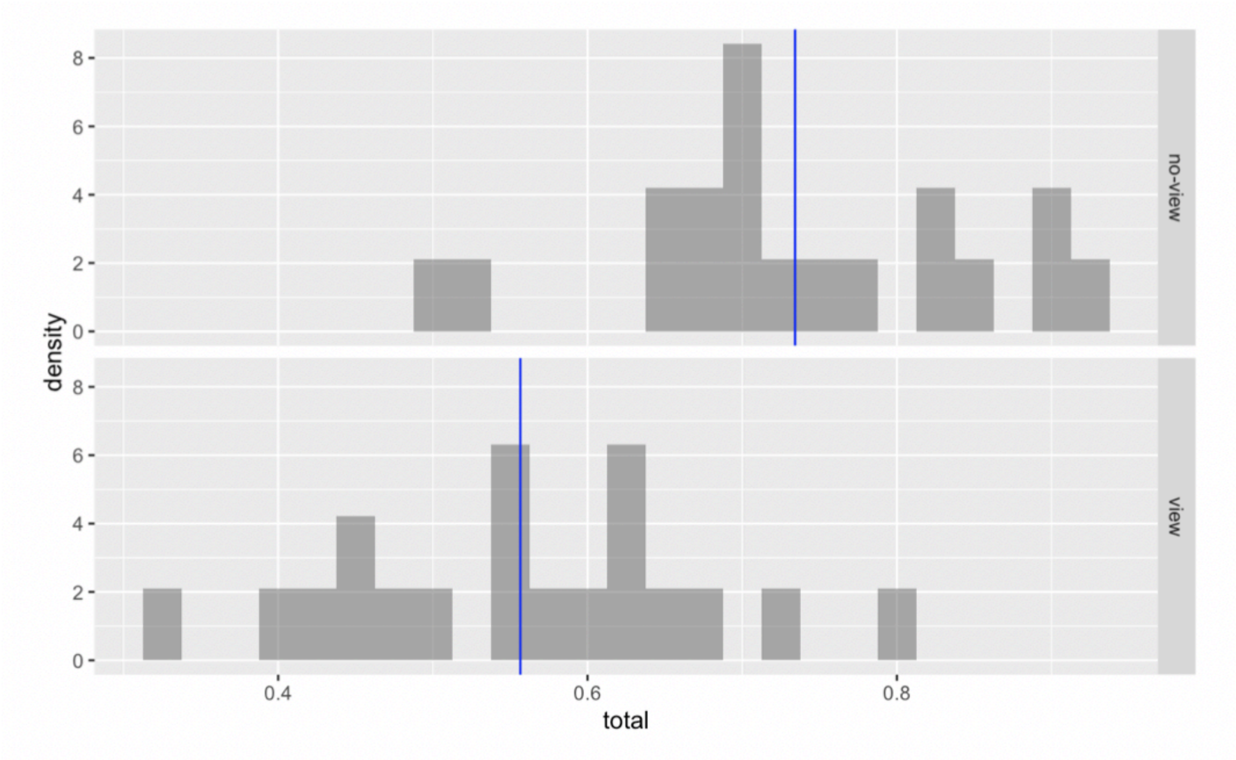
3. What would happen to the number of observations in the view condition?
    a. The number of observations would increase
    b. The number of observations would stay the same
    c. The number of observations would decrease
    d. The number of observations would increase, decrease, or stay the same, but it's impossible to tell which

4. Explain your answer to the previous question

We used the code below to create a faceted histogram showing the distribution of total in each condition. The vertical lines represent mean total scores for the two conditions. Again, you can see that the participants in the no-view group scored higher, on average, than participants in the view group.

```
stats <- favstats(total ~ condition, data = laptop_data)

gf_dhistogram(~ total, data = laptop_data) %>%
 gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
 gf_facet_grid(condition ~ .)
```



5. Sometimes groups differ just because of randomness. Do you think the group difference in the histogram above could be due to randomness?
    a. Yes, it must be due to randomness
    b. No, it cannot be due to randomness
    c. Maybe, need to further investigate

6. Explain your answer to the previous question:

7. If you wanted to investigate whether this difference could be due to randomness using the shuffle() function, what would you do?
    Please be as specific as possible in your response.

8. Alex thinks she only needs to shuffle **once** to see if the difference between conditions on total could be due to randomness by comparing the shuffled result with the original data. Mary thinks she needs to shuffle more than once to be able to see if the difference could be due to randomness. Do you agree with Alex or Mary? Explain your answer.

Take a look at each line of code below. For each line, **explain 1) what the code is doing** and **2) why someone would write that code**.

```
laptop_data$condition.shuffle <- shuffle(laptop_data$condition)
```

9. What is this line of code doing?

10. Why would someone write this line of code?

```
laptop_data$total.shuffle<- shuffle(laptop_data$total)
```

11. What is this line of code doing?

12. Why would someone write this line of code?

A Look at the two examples of codes below. Example 1 and Example 2 each produces a faceted histogram.

Example 1:

```
gf_dhistogram(~ distracted , data = laptop_data) %>%
gf_facet_grid(shuffle(condition) ~ .)
```

Example 2:

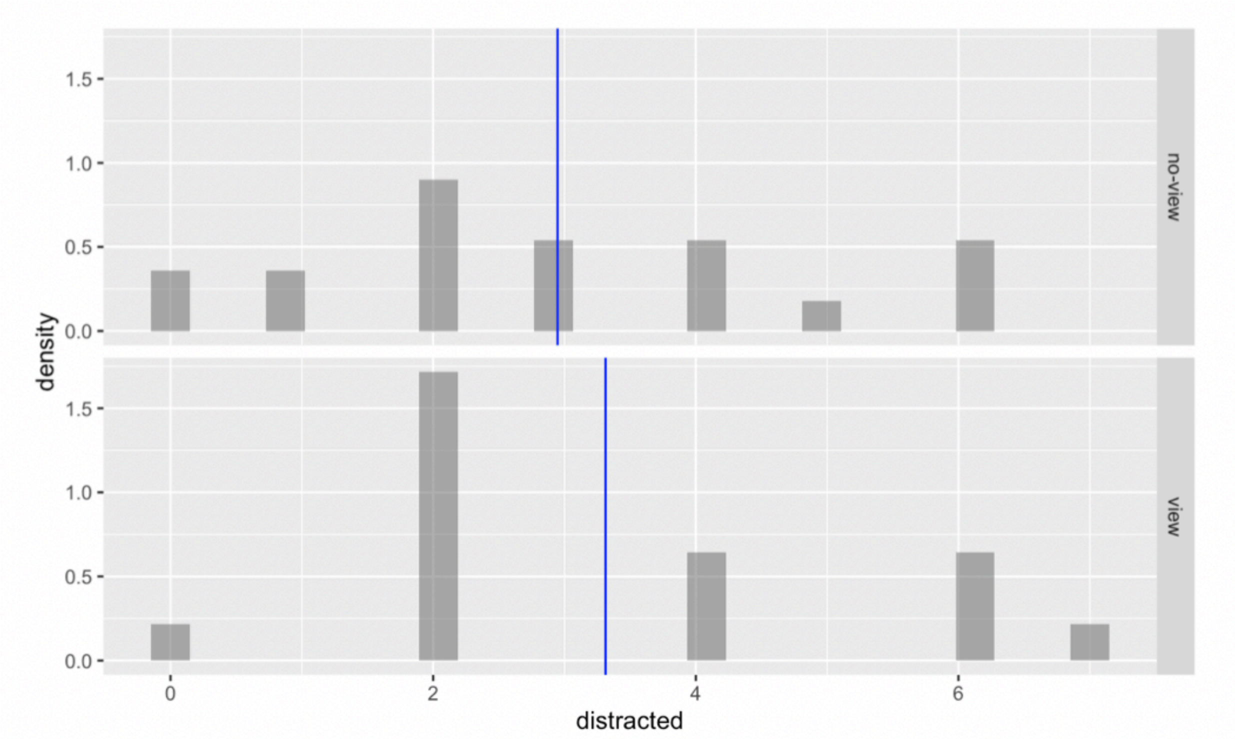
```
gf_dhistogram(~ shuffle(distracted) , data = laptop_data) %>%
gf_facet_grid(shuffle(condition) ~ .)
```

13. In what ways would the two faceted histograms be similar?

14. In what ways would the two faceted histograms be different?

We ran this code to create the graph below. We added a line in each condition to represent the mean of **distracted** of that **condition**. Notice that the average **distracted** rating in the **no-view condition** is lower than the average **distracted** rating in the **view condition**.

```
stats <- favstats(distracted ~ condition, data = laptop_data)
gf_dhistogram(~ distracted, data = laptop_data) %>%
gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
gf_facet_grid(condition ~ .)
```



15. Sometimes groups differ just because of randomness. Do you think the group difference in the histogram above could be due to randomness?
    a. Yes, it must be due to randomness
    b. No, it cannot be due to randomness
    c. Maybe, we need to further investigate

16. Explain your answer to the previous question:

17. If you ran the code in the previous question again, do you think it would produce the same output?
    a. Yes
    b. No
    c. It's possible, but not likely

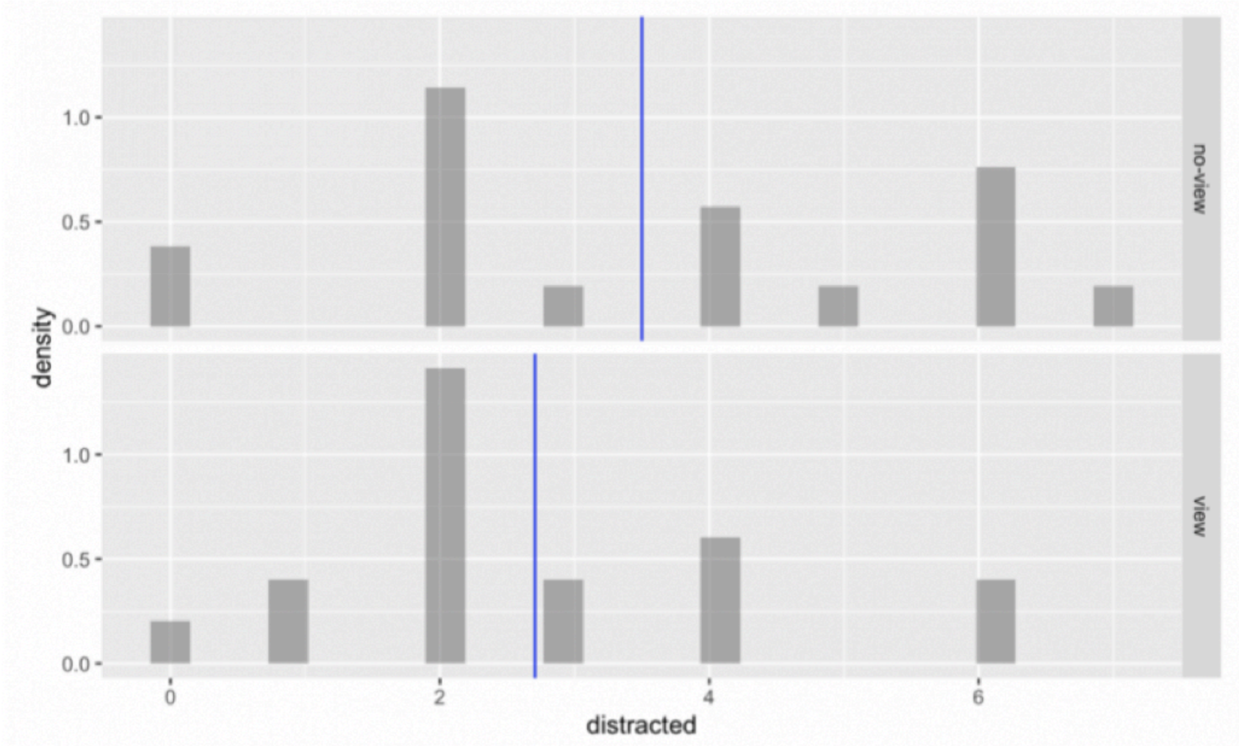
18. Explain your answer to the previous question:

We revised the code from the previous question to create the graph below. We added a line to represent the mean of **distracted** for each **condition**. Notice that the average **distracted** rating in the **no-view condition** is higher than the average **distracted** rating in the **view condition**.

```
laptop_data$condition.shuffle <- shuffle(laptop_data$condition)
```

```
stats <- favstats(distracted ~ condition.shuffle, data = laptop_data)

gf_dhistogram(~distracted, data = laptop_data) %>%
 gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
 gf_facet_grid(condition.shuffle ~ .)
```



19. What caused the difference in the means represented in the graphs below?

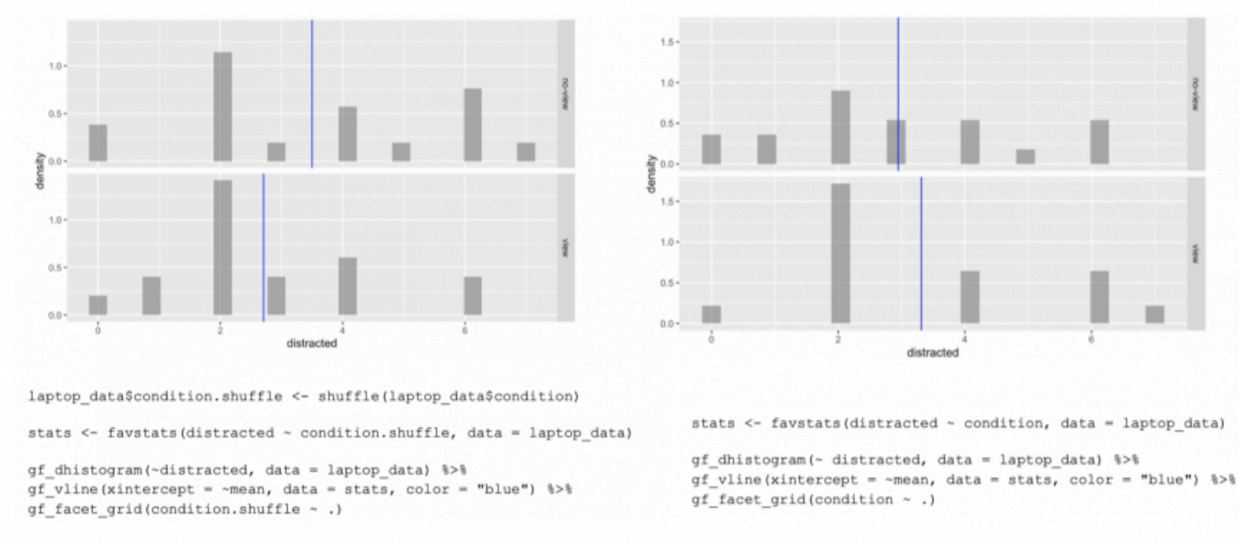
20. Sometimes groups differ just because of randomness. Do you think the group difference in the histogram above could be due to randomness?
    a. Yes, it must be due to randomness
    b. No, it cannot be due to randomness
    c. Maybe, need to further investigate

21. Explain your answer to the previous question:

22. If you ran the code in the previous question again, do you think it would produce the same output?
    a. Yes
    b. No
    c. It's possible, but not likely

23. Explain your answer to the previous question

Look at the two faceted histograms below, along with the code that produced each (the code might be a bit hard to read, feel free to zoom in to get a better read):



24. Why do the two faceted histograms look different?

25. Based on what you've learned from these two histograms, do you think being able to view or not view a laptop during class (condition) affects students' self-reported rating of how distracted they were in class (as measured by distracted score on a post-lesson assessment)? Why or why not?

Imagine we run the code below:

```
laptop_data$distracted.shuffle <- shuffle(laptop_data$distracted)

mean(laptop_data$distracted.shuffle)
mean(laptop_data$distracted)
```

26. How would the mean of distracted.shuffle compare to the mean of distracted?
    a. The mean of distracted.shuffle would be larger
    b. The mean of distracted.shuffle would be smaller
    c. The two means would be the same
    d. It's impossible to tell

27. Explain your answer to the previous question:

28. What will the distribution of **distracted.shuffle** look like compared to the distribution of **distracted**?
    a. Wider
    b. Narrower
    c. The same
    d. Not sure. It will vary randomly.

29. Explain your answer to the previous question:

Post-Test Attitudinal Measures

1. How well do you understand what the shuffle() function does? (with 0 being not at all)

Please rate your level of agreement with each of the following statements:

2. I learned a lot from this activity
    a. Strongly agree
    b. Agree
    c. Somewhat agree
    d. Neither agree nor disagree
    e. Somewhat disagree
    f. Disagree
    g. Strongly disagree

3. I like this way of learning R functions
    a. Strongly agree
    b. Agree
    c. Somewhat agree
    d. Neither agree nor disagree
    e. Somewhat disagree
    f. Disagree
    g. Strongly disagree

4. On a scale of 1 to 6 (with 1 being not at all confident and 6 being extremely confident), how confident do you feel in your R skills?

References

Acevedo Nistal, A., Van Dooren, W., Clarebout, G., Elen, J., & Verschaffel, L. (2009). Representational fluency and flexibility in the domain of linear functions: A choice/no-choice study. In *Fostering Communities of Learners. Biennial Conference of the European Association for Research on Learning and Instruction, Date: 2009/08/25-2009/08/29, Location: Amsterdam, The Netherlands*.

Ainsworth, S., & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and instruction*, *14*(3), 241-255. https://doi.org/10.1016/j.learninstruc.2004.06.002

Alibali, M. W., Flevares, L. M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: do teachers have the upper hand? *Journal of Educational Psychology, 89*, 183–193. doi: 10.1037/0022-0663.89.1.183. https://doi.org/10.1037/0022-0663.89.1.183

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. doi: 10.1177/1088868309341564

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559. https://doi.org/10.1126/science.1736359

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deiectic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*, 723-767.

Chance, B., del Mas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Springer, Dordrecht. https://link.springer.com/book/10.1007/1-4020-2278-6#toc

Chance, B., & Rossman, A. (2006, July). Using simulation to teach and learn statistics. In *Proceedings of the Seventh International Conference on Teaching Statistics* (pp. 1-6). Voorburg, The Netherlands: International Statistical Institute.

Cheng, P. C. H. (1999). Unlocking conceptual learning in mathematics and science with effective representational systems. *Computers & Education*, *33*(2-3), 109-130. https://doi.org/10.1016/S0360-1315(99)00028-7

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, *143*(2), 694-709. http://doi.org/10.1037/a0033861

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child development*, *84*(6), 1863-1871. https://doi.org/10.1111/cdev.12097

Da Rold, F. (2018). Defining embodied cognition: The problem of situatedness. *New Ideas in Psychology*, *51*, 9-14. https://doi.org/10.1016/j.newideapsych.2018.04.001

delMas, R. C., Garfield, J., & Chance, B. (1999). A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning, *Journal of Statistics Education, 7*(3).

Dyck, J. L., & Gee, N. R. (1998). "A Sweet Way to Teach Students About the Sampling Distribution of the Mean," *Teaching of Psychology*, 25, 192–195. https://doi.org/10.1207/s15328023top2503_6

Fries, L., Son, J. Y., Givvin, K. B., & Stigler, J. W. (2021). Practicing connections: A framework to guide instructional design for developing understanding in complex domains.

Educational Psychology Review, *33*(2), 739-762. https://doi.org/10.1007/s10648-020-09561-x

Fu, Y., & Franz, E. A. (2014). Viewer perspective in the mirroring of actions. *Experimental Brain Research,* 232(11): 3665–3674. https://doi.org/10.1007/s00221-014-4042-6

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. https://doi.org/10.1007/s10648-014-9249-3

Fyfe, E. R., & Nathan, M. J. (2019). Making "concreteness fading" more concrete as a theory of instruction for promoting transfer. *Educational Review*, *71*(4), 403-422.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability10. *SERJ Editorial Board*, 92.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of educational psychology, 96*(3), 424. https://doi.org/10.1037/0022-0663.96.3.424

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology, 64*: 257–283. https://doi.org/10.1146/annurev-psych-113011-143802

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological science*, *12*(6), 516-522.

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: using children's hands to read their minds. *Cognition and Instruction, 9*, 201–219. https://doi.org/10.1207/s1532690xci0903_2

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the learning sciences, 14*(1), 69-110. https://doi.org/10.1207/s15327809jls1401_4

Hancock, S. A., & Rummerfield, W. (2020). Simulation methods for teaching sampling distributions: Should hands-on activities precede the computer? *Journal of Statistics Education*, *28*(1), 9-17. https://doi.org/10.1080/10691898.2020.1720551

Hodgson, T., & Burke, M. (2000). On simulation and the teaching of statistics. Teaching Statistics, 22(3), 91-96. http://doi.org/10.1111/1467-9639.00033

Kastens, K. A., Liben, L. S., & Agrawal, S. (2008). *Epistemic actions in science education*. Basel: Proceedings of Spatial Cognition. https://doi.org/10.1111/1467-9639.00033

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & Jupyter Development Team. (2016). *Jupyter Notebooks-a publishing format for reproducible computational workflows* (Vol. 2016, pp. 87-90). https://doi.org/10.3233/978-1-61499-649-1-87

Kokkonen, T., & Schalk, L. (2021). One Instructional sequence fits all? A conceptual analysis of the applicability of concreteness fading in mathematics, physics, chemistry, and biology education. *Educational Psychology Review*, *33*(3), 797-821.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of experimental psychology: General*, *126*(4), 349. https://doi.org/10.1037/0096-3445.126.4.349

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory

judgments. *Psychological Science*, *22*(6), 787–794.
https://doi.org/10.1177/0956797611407929

Lane, D. M. (2015). "Simulations of the Sampling Distribution of the Mean Do Not Necessarily Mislead and Can Facilitate Learning," *Journal of Statistics Education*, 23. doi: https://doi.org/10.1080/10691898.2015.11889738

Lunsford, M. L., Rowell, G. H., & Goodson-Espy, T. (2006). Classroom research: Assessment of student understanding of sampling distributions of means and the central limit theorem in post-calculus probability and statistics classes. *Journal of Statistics Education*, *14*(3). https://doi.org/10.1080/10691898.2006.11910587

Paas, F., & van Merriënboer, J. J. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science, 29*(4), 394-398. https://doi.org/10.1177/0963721420922183

Pfaff, T. J., & Weinberg, A. (2009). Do hands-on activities increase student understanding?: A case study. *Journal of Statistics Education*, *17*(3). https://doi.org/10.1080/10691898.2009.11889536

Pouw, W. T., Van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*(1), 51-72.

Pruim, R., Kaplan, D. T., & Horton, N. J. (2017). The mosaic Package: Helping Students to Think with Data Using R. *R J.*, *9*(1), 77. https://doi.org/10.32614/RJ-2017-024

Rueckert, L., Church, R. B., Avila, A., & Trejo, T. (2017). Gesture enhances learning of a complex statistical concept. *Cognitive Research: Principles and Implications, 2*(1), 2. https://doi.org/10.1186/s41235-016- 0036-1

Savinainen, A., Scott, P., & Viiri, J. (2005). Using a bridging representation and social interactions to foster conceptual change: Designing and evaluating an instructional sequence for Newton's third law. *Science Education*, *89*(2), 175-195. https://doi.org/10.1002/sce.20037

Sepp, S., Howard, S. J., Tindall-Ford, S., Agostinho, S., & Paas, F. (2019). Cognitive Load Theory and Human Movement: Towards an Integrated Model of Working Memory. *Educational Psychology Review, 31*(2), 293-317. https://doi.org/10.1007/s10648-019-09461-9

Son, J. Y., Ramos, P., DeWolf, M., Loftus, W., & Stigler, J. W. (2018). Exploring the practicing-connections hypothesis: using gesture to support coordination of ideas in understanding a complex statistical concept. *Cognitive research: principles and implications, 3*(1). https://doi.org/10.1186/s41235-017-0085-0.

Spiro, R. J. (1988). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. *Center for the Study of Reading Technical Report; no. 441*. https://doi.org/10.1598/0710.22

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review, 22*(2), 123-138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development, 68*(1), 1-16. https://doi.org/10.1007/s11423-019-09701-3

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261-292. https://doi.org/10.1007/s10648-019-09465-5

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of educational psychology, 80*(4), 424. https://doi.org/10.1037/0022-0663.80.4.424

Tran, C., Smith, B., & Buschkuehl, M. (2017). Support of mathematical thinking through embodied cognition: Nondigital and digital approaches. *Cognitive Research: Principles and Implications, 2*(1), 1-18. https://doi.org/10.1186/s41235-017-0053-8

Varga, S., & Heck, D. H. (2017). Rhythms of the body, rhythms of the brain: Respiration, neural oscillations, and embodied cognition. *Consciousness and Cognition*, *56*, 77-90.

Watkins, A. E., Bargagliotti, A., & Franklin, C. (2014). "Simulation of the Sampling Distribution of the Mean Can Mislead," *Journal of Statistics Education*, 22. https://doi.org/10.1080/10691898.2014.11889716.

Weisberg, S. M., & Newcombe, N. S. (2017). Embodied cognition and STEM learning: Overview of a topical collection in CR: PI. *Cognitive Research: Principles and Implications*, *2*(1), 1-6. https://doi.org/10.1186/s41235-017-0071-6

Wood, M. (2005). "The Role of Simulation Approaches in Statistics," Journal of Statistics Education, [Online], 13 (3). https://doi.org/10.1080/10691898.2005.11910562

Zhang, X., & Maas, Z. (2019). Using R as a Simulation Tool in Teaching Introductory Statistics. *International Electronic Journal of Mathematics Education*, *14*(3), 599-610. https://doi.org/10.29333/iejme/5773

Zieffler, A., Garfield, J., Delmas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2). https://iase-web.org/documents/SERJ/SERJ7(2)_Zieffler.pdf?1402525008

Chapter 2

Watching Hands Move Enhances Learning from Concrete and Dynamic Visualizations

Icy Zhang, Ji Y. Son, and James W. Stigler

University of , Los Angeles

Abstract

This article explores the role of sensorimotor engagement in students' learning of a challenging STEM-related concept. Previous research has failed to distinguish two features commonly associated with embodiment: sensorimotor engagement and visuospatial concreteness. In the current research, we ask whether sensorimotor engagement – operationalized as watching a video of hands manipulating paper representations – offers unique benefits beyond the visuospatial concreteness of a dynamic visualization of the same process. Participants were randomly assigned to one of three conditions to learn about the shuffle() function in R: a Watch Hands Moving Objects (WHMO) group, which watched a video with hands; a Watch Moving Objects (WMO) group, which watched a video with a dynamic visualization in which objects moved without hands; or a control group, which watched a live-coding video that did not include either hands or visuospatial representations. Results revealed that only participants in the WHMO group demonstrated significantly superior performance compared to both the WMO group and control groups. These findings highlight the unique benefit of sensorimotor engagement for learning, contributing to a deeper understanding of how embodiment can enhance the learning process.

Keywords: embodied cognition, multimedia learning, dynamic visualization, sensorimotor engagement, data science education

Public Significance Statement

This research investigates the unique contribution of sensorimotor engagement when learning from instructional videos. Students who watched an instructional video in which hands were seen moving objects representing a statistical programming concept learned more than those who watched a video that included the same dynamic visualization but without the hands. The work contributes to the theory of embodied learning and suggests a simple way to improve educational materials.

Watching Hands Move Enhances Learning from Concrete and Dynamic Visualizations

In recent decades, there has been a significant increase in research on embodied cognition, especially in education, where a multitude of studies have demonstrated that learning can be enhanced through various forms of embodiment (see Shapiro & Stolz, 2019 for a review). Both observing and performing bodily movements have been shown in a number of studies to aid student learning (Cook et al., 2016; Goldin-Meadow et al., 2001; Johnson-Glenberg et al., 2014, 2016; Ping & Goldin-Meadow, 2008; Pouw et al., 2016; Tran et al., 2017; Zhang et al., 2022). This benefit has been demonstrated in a variety of STEM fields, including mathematics (Nathan & Alibali, 2011; Novack & Goldin-Meadow, 2015), data science (Zhang et al., 2022), and physics (Johnson-Glenberg et al., 2014; Johnson-Glenberg & Megowan-Romanowicz, 2017).

Although the beneficial effects of embodiment on learning have been extensively studied, past research has typically compared the effect of embodied versus non-embodied approaches without further isolating different features of embodiment. Here, we identify at least three important features of embodiment to consider: sensorimotor engagement, visuospatial concreteness, and dynamic quality. *Sensorimotor engagement* refers to the level of involvement of the sensorimotor system during learning. It can range from physically interacting with objects to imitating actions to simply observing someone's gestures or object manipulations. *Visuospatial concreteness* refers to the degree to which a concept or stimulus can be experienced in a material or objectified manner, representing abstract concepts through concrete representations. Lastly, *dynamic quality* characterizes the degree to which perceptual stimuli move or change during the instruction.

One of the most examined fields of study in embodied learning compares the effect of more embodied learning experiences or materials that are high on all three of these features against experiences that are lower on these features. For example, Goldin-Meadow et al.'s (2001) finding that simply allowing learners to gesture during learning improved learning outcomes was based on a comparison of a condition that included more sensorimotor engagement, more visual concreteness, and more movement (or dynamic quality) to one that was low on all three features (i.e., not allowing learners to gesture). Similarly, Johnson-Glenberg et al.'s (2014) finding that moving the learners' entire upper body during instruction benefits learning in science domains was based on a comparison of a condition that included more sensorimotor engagement, more visual concreteness, and more movement (or dynamic quality) to one that was low on all three features (i.e., a regular instruction).

The field of embodied cognition is just beginning to investigate how these features might interact during learning. Learning experiences that involve the body or hands (i.e., high on sensorimotor engagement) are almost always dynamic but they can vary in their degree of visuospatial concreteness. For example, object manipulation and gesture are two dynamic hand actions commonly leveraged in science education to benefit learners (Yammine & Violato, 2016; Roberts et al., 2005; Novack & Goldin-Meadow, 2015). They are both high on sensorimotor engagement, but object manipulation has relatively more visuospatial concreteness (Castro-Alonso et al., 2019). Because object manipulation requires both the movement of hands and concrete objects, it has higher visuospatial concreteness than gesture alone (Chu & Kita, 2008).

Evidence suggests that people interpret hand movements differently in the presence and absence of objects (Schachner & Carey, 2013). When objects are absent, hand movements are interpreted in terms of movement-based goals and when objects are present, they are interpreted relative to external goals (Novack et al., 2016). However, in at least one study, such differences do not have implications for learning: gestures can benefit learners both in the presence and absence of objects (Ping & Goldin-Meadow, 2008). This study can be construed as evidence that a dynamic learning situation with *sensorimotor* features alone was as beneficial as one with *sensorimotor+concrete* features. But more research to disentangle these features is needed.

Similarly, dynamic learning experiences can also be high in visuospatial concreteness but low in sensorimotor engagement (e.g., squares moving around in space on their own) or they can be high in both dimensions (e.g., hands moving square objects around). One study found that people learn better from dynamic drawing with a visible hand than from already drawn diagrams (Fiorella & Mayer, 2016). In this case, a *sensorimotor+concrete+dynamic* presentation of a diagram was better than one that was merely visuospatially *concrete*. Although studies abound that show the benefits of performing actions as well as watching the actions of others (for a review, see Goldin-Meadow & Beilock, 2010), studies of embodied learning interventions, to our knowledge, have not directly examined whether sensorimotor engagement has any added benefits above and beyond that of visuospatial concreteness. That is, how would a *sensorimotor+concrete+dynamic* presentation compare to one that is *concrete+dynamic*? Such evidence is crucial for advancing the field, and thus is the focus of our investigation.

One particularly relevant vein of research provides insight to this question by comparing performing *versus* observing actions. For example, Goldin-Meadow and colleagues (2012) compared the effects of performing versus observing gestures on a mental rotation task that asked six-year-olds to judge whether two shapes at different angles of rotation were the same or different. Children who were instructed to perform gestures that mimicked the rotation of the figures performed better on the task than children who simply observed someone else performing similar gestures on a video clip. Performing gestures involves higher sensorimotor engagement than observing them, but has the same level of visuospatial concreteness and dynamic quality. Goldin-Meadow et al. have shown a clear effect of a high level of sensorimotor engagement vs. a medium level of engagement (observing), but leaves open the question of whether observing sensorimotor activity benefits learning above a condition with no observable sensorimotor activity.

With dynamic learning stimuli such as those found in instructional videos, it is difficult to disentangle the role of sensorimotor engagement from visuospatial concreteness. Our approach compares an embodied intervention, operationalized as an instructional video with both sensorimotor engagement and visuospatial concreteness, against a similar video that preserves the visuospatial concreteness but reduces sensorimotor engagement. We designed dynamic instructional videos to teach students learning about randomness in statistics how to use the shuffle function (from the programming language R) which randomly re-orders individual elements (either rows or cells in a data set). In the embodied intervention, students are shown a video of hands physically cutting up a dataset printed on paper and reordering cells. This intervention involves both sensorimotor engagement, through the instructor's hand movements, and visuospatial concreteness, through physical paper objects and spatial manipulation of those objects. In the less embodied version, the video shows the data set being split up and re-ordered without the involvement of any hands. The objects simply move on their own.

The current research aims to investigate whether sensorimotor engagement, such as watching the hands physically manipulate paper, offers unique benefits beyond representational concreteness in dynamic learning videos. After all, concrete representations may benefit learning without a sensorimotor component. For example, the moving data frames and cells may facilitate learning by associating abstract ideas (such as "variables" and "values") with more concrete visuospatial objects (such as columns and cells, respectively). Such representations can implicitly represent abstract properties in an analog fashion (Goldstone & Barsalou, 1998) without the need for explicit statements and memorization of assumptions (e.g., when cells of a data frame are being moved around in space, students can see that cells are not being added or taken away). Prior research has shown that concrete objects are particularly beneficial for young children and novices (Fyfe et al., 2014; Montessori, 1917; Piaget, 1970; Uttal et al., 2006) presumably because they have less background knowledge and need to use concrete features to build up a basis for understanding new abstract concepts. This has led to the concreteness-fading hypothesis, which hypothesizes that instruction should transition from concrete to abstract for optimal learning (Fyfe et al., 2014). Many studies have demonstrated the efficacy of such an instructional sequence, demonstrating the benefits of visuospatial concreteness early in the learning process (e.g. Fyfe et al., 2015).

Beyond concreteness, are there additional learning benefits of engaging the sensorimotor system in a dynamic learning video? If sensorimotor engagement is not found to add additional value beyond visuospatial concreteness in dynamic learning stimuli, it may challenge a fundamental tenet of embodied cognition, that the body plays a unique role in cognitive activities. Therefore, the present study aims to disentangle the effects of visuospatial concreteness and sensorimotor engagement in embodied learning research. Compared to more abstract instruction, does concreteness *per se* lead to better learning or do we need both concreteness and sensorimotor engagement to see benefits to learning?

Current Study

The current study is based on a previous study (Zhang et al., 2022) in which college students were randomly assigned to watch an instructional video featuring either hands-on demonstrations or live-coding in R prior to watching a second live-coding video using a larger data set. The results showed that students who watched a hands-on demonstration before watching a live-coding video learned more than those who watched two live-coding videos in a row.

However, the hands-on video used in that study involved both visuospatial concreteness (i.e., pieces of paper to represent data) and sensorimotor engagement (i.e., hands shuffling the paper). In the current study, we will refer to this as the *Watch Hands Moving Objects* (*WHMO*) condition. We also will introduce a new condition that has visuospatial concreteness without any hand movements, which we will refer to as the *Watch Moving Objects* (*WMO*) condition. In this condition students saw visuospatial representations of the dataset and cells moving around dynamically, but without being manipulated by the instructor's hands. Although this condition might still involve some sensorimotor engagement (via the presentation of visuospatial objects), it is at a lower level than the hands-on video, making it suitable for addressing our research question. We also included Zhang et al.'s (2022) control condition, which showed the same concepts being taught through live coding alone. Table 1 summarizes how the three conditions vary on three key features: sensorimotor engagement, visuospatial concreteness, and dynamic quality.

The embodied cognition view would be that sensorimotor engagement confers a unique benefit to learning. This view would expect that the type of instructional video students are exposed to before the abstract live-coding video would significantly impact their subsequent learning outcomes. In the current study specifically, we hypothesize that students who first watch a hands-on video, which involves sensorimotor engagement, visuospatial concreteness, and dynamic qualities will perform better than those who watch videos with only visuospatial concreteness and dynamic qualities, or the control (i.e., the live-coding).

Table 1

Summary of the Three Conditions Based on the Three Features

Condition	Sensorimotor Engagement	Visuospatial Concreteness	Dynamic Quality
Control (live-coding)	Low	Low	Medium
WMO (Watch Moving Objects)	Low	High	High
WHMO (Watch Hands Moving Objects)	High	High	High

Beyond exploring whether sensorimotor engagement leads to better learning, we also wish to understand the potential mechanisms by which embodiment causes better learning. What are the mental representations that result from a more embodied learning experience? To explore this question, at the end of our study we will ask participants to describe whether they thought about the contents of the video while answering the post-test questions and if so, what they thought about. Specifically, we are interested in whether their recalls were visuospatial (e.g., recall learning that data frames are made up of rows, columns, and cells, and that R code can change the arrangement of these elements.

If watching sensorimotor activity changes the quality and content of mental representations, as several studies in gesture have demonstrated (e.g., Alibali et al., 2000; Brooks et al., 2018; Rimé et al., 1984, Wagner et al., 2004), participants in the WHMO condition should stand apart from those in the other two conditions in both the quantity and quality of their recall of the learning videos. For example, they may be more likely to recall that cells are "moving," even though the cells are moving in both the WHMO and WMO conditions. If this were the case, we would have evidence that sensorimotor engagement even by watching sensorimotor activities uniquely leads to different mental representations.

If, on the other hand, the quantity and quality of recall look similar across the WHMO and WMO conditions, yet different from the control condition, we might conclude that it is the dynamic and concrete qualities of the representations that impacts learning, and not the sensorimotor engagement.

For the quality of their recall, we asked specifically whether the elements recalled were visuospatial. This is an important question because past research has revealed mixed evidence of whether the effect of gesture is visuospatial or propositional (Alibali et al., 2000; Wagner et al., 2004). If we observe participants reporting more visuospatial recall in the WHMO condition, this might provide support for a visuospatial representation that underlies embodiment. However, if we only see differences in general recall, but no difference when it comes specifically to visuospatial recall between the WHMO condition and the other two conditions, we might lean toward a propositional representation.

Method

Participants

Participants were 153 undergraduate students taking an introductory psychological statistics course at a large public research institution. They took the course either during the summer of 2022 or the winter of 2023. These students were selected to participate in the study because they had already been introduced to the shuffle function from the mosaic package in R (v1.8.4; Pruim et al., 2017) as part of their coursework. Mastery of the shuffle function, which instantiates the process of randomness, was integral to understanding further course topics, such as the sampling distribution.

Three participants were excluded because they took more than 5 hours on the study survey, indicating that they did not complete the study in one sitting as required. This resulted in a final sample size of 150 students.The gender and ethnic composition of the sample matched that of the course. The sample contained 120 females and 30 males. The racial/ethnic breakdown was as follows: 64 Asian, 6 Black or African American, 1 Native Hawaiian or other Pacific Islander, 32 other or mixed race, and 47 White. This information was collected through self-reports. Students who participated in the study received a small amount of extra credit (0.5%) toward their course grades and also may have derived educational benefits from their participation. The study was reviewed and approved by the university's institutional review board.

A power analysis was conducted using the pwr package in R (Champely et al., 2017; Cohen, 1988). Based on a Cohen's f of 0.3, with an α of .05 and a power of .85, the minimum sample size needed with this effect size is 42 per group. The power analysis indicated that the sample is sufficiently large to detect a medium effect size.

Design & Procedure

The study was hosted on Qualtrics and students participated online. Participants received an email from their professor with the Qualtrics link to the study. They volunteered to participate by clicking on the link, upon which Qualtrics randomly assigned participants to one of the three conditions: *Control* (n = 47), *WMO* (n = 52), and *WHMO* (n = 51).

After answering some basic demographic questions, participants responded to five open-response questions designed to assess their existing understanding of probability and the shuffle function. Subsequently, they viewed two intervention videos, detailed below in the materials section. The first video varied according to experimental condition, while the second video was identical across the three conditions. After watching the second video, participants answered 22 post-test questions. They then answered two additional questions about how often they thought about the content of the videos during the time they were answering the posttest questions, and, if they did think at all about the videos, what they were thinking about specifically.

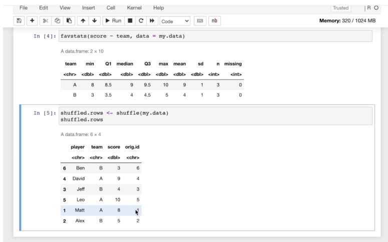
Intervention Videos

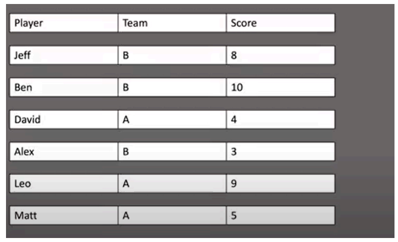
Students started by watching one of three versions of an instructional video. It is worth highlighting that although some students watched videos that included hands-on activities, participants in this study did not perform any hands-on activities themselves, a fact that is reflected in our naming of the conditions (the “W” stands for “Watching”). Although the format differed, the content of the videos was carefully matched across versions. All three versions of the video explained the use of the shuffle function to simulate randomness by showing what a small artifactual dataset would look like after shuffling rows versus cells within columns. A screenshot from each video is shown in Figure 4.

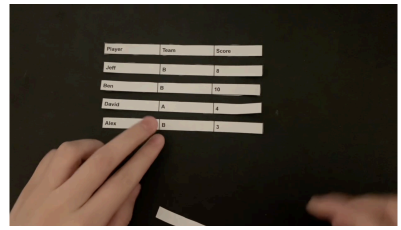
Figure 4

Screenshots of Each Video from the Three Conditions

Control	WMO	WHMO







Students in the *WHMO* (Watch Hands Moving Objects) condition watched a video, shot from above, of the instructor's hands as she cut a printed data table into pieces and then randomly rearranged the pieces. In other words, the pieces of the datasets were cut and moved by the instructors' hands. This approach provided a concrete and sensorimotor representation of what the shuffle function does. As the pieces of data were manually shuffled, the narrator explained what was being done.

Students in the *WMO* (Watch Moving Objects) condition heard the same narration as those in the *WHMO* condition, but instead of watching a person cut and shuffle pieces of paper, they saw animated visualizations of the data table being separated and shuffled. The animations, which were generated by PowerPoint, were matched with the movements of the physical dataset in the *WHMO* video but, critically, did not show any hands manipulating the data set. Our goal was to simulate the WHMO experience but without the hands, thus minimizing the activation of sensorimotor systems.

Students assigned to the control condition saw a computer screen recording of R code being typed and executed to perform the same shuffles as those enacted in the WHMO and WMO videos. The narration was the same, except that it referred to the code being run instead of the visuospatial movements described in the other two videos. The control condition provided a less perceptually concrete experience because the pieces of the datasets were not being cut out or moved around as in the other two conditions. The data set would simply appear or change after the code was run. However, it is worth noting that this condition was also dynamic because of the live coding - it was just less dynamic than the two experimental conditions. It was not entirely abstract either, as the instructor wrote code to print out the dataset before and after running the shuffle function to show changes in the data set, thereby retaining a degree of visuospatial concreteness in the learning experience.

After watching their assigned version of the first video, students in all three conditions watched a live-coding video, which was similar in format to the live-coding video described above. In this second video, the live coding involved applying concepts learned in the first video to a larger dataset adapted from a real experiment. The descriptions of the WHMO condition and the control condition have been published previously in Zhang et al. (2022).

Measures

Pretest

Participants' knowledge before watching the videos was measured on a pretest consisting of five open-response questions (Appendix A).

Post-test

Participants' knowledge after watching the videos was measured using a combination of eight multiple-choice questions and 25 open-response questions. The questions were designed to evaluate students' understanding of the shuffle function, the concept of randomness, and how to use the concept of randomness to make statistical inferences. (A complete list of the questions is presented in Appendix B.)

Both pre- and post-test questions were graded by two trained coders based on a predetermined rubric. The two coders were blind as to the experimental condition from which each response came. Each question was given a maximum of one point. Partial credit of 0.5 was given to open responses that were missing parts or showed minor misunderstandings. The possible total score ranged from 0 to 32 (Cronbach's α = .89).

Recall of Instructional Videos during Problem-Solving

General Recollection of Video Content. After completing the posttest questions, participants were asked: "When you were answering the posttest questions, how often did you think about the content in the videos?" Responses were coded on a five point scale based on a predetermined rubric (see Table 2). After all responses were coded by one experimenter, another trained experimenter coded 20% of the responses to establish interrater reliability (Cronbach's α = .98).

Table 2

Coding Rubric for Reflection of Video Content

Score	Coding
4	all the time/ every question
3	often/ half of the questions
2	sometimes/ a couple questions
1	Not often/ only one or two questions
0	Not at all/ never/ none of the questions

Visual Recall of the video. A follow-up question asked, "If you did think about the video, what did you think about specifically?" A trained experimenter coded whether participants referenced visual elements from the video or not. For example, mentions of specific actions like "cutting up paper and 'shuffling' data" or recalling distinct images such as histograms from the video were dummy coded with a 1, signifying visual recall. In contrast, references to abstract concepts or non-visual elements, like the idea of "shuffling to break the relationship" or the importance of "running a function multiple times to explain variation," were coded as 0, no visual recall. A second experimenter coded 20% of the responses to this question (Cronbach's α = .86).

Transparency and openness

The way we determined our sample size, excluded participants, all manipulations and measures followed the Journal Article Reporting Standards (Kazak, 2018). All data is available at https://osf.io/y795h/?view_only=c8c8ad03fe74462397e85ef84708e4d3. Analysis code, and research materials are available upon request. Data were analyzed using R, version 4.0.0 (R Core Team, 2020) and data visualizations using the package ggplot2, version 3.2.1 (Wickham, 2016). Neither the study's design nor its analysis were pre-registered.

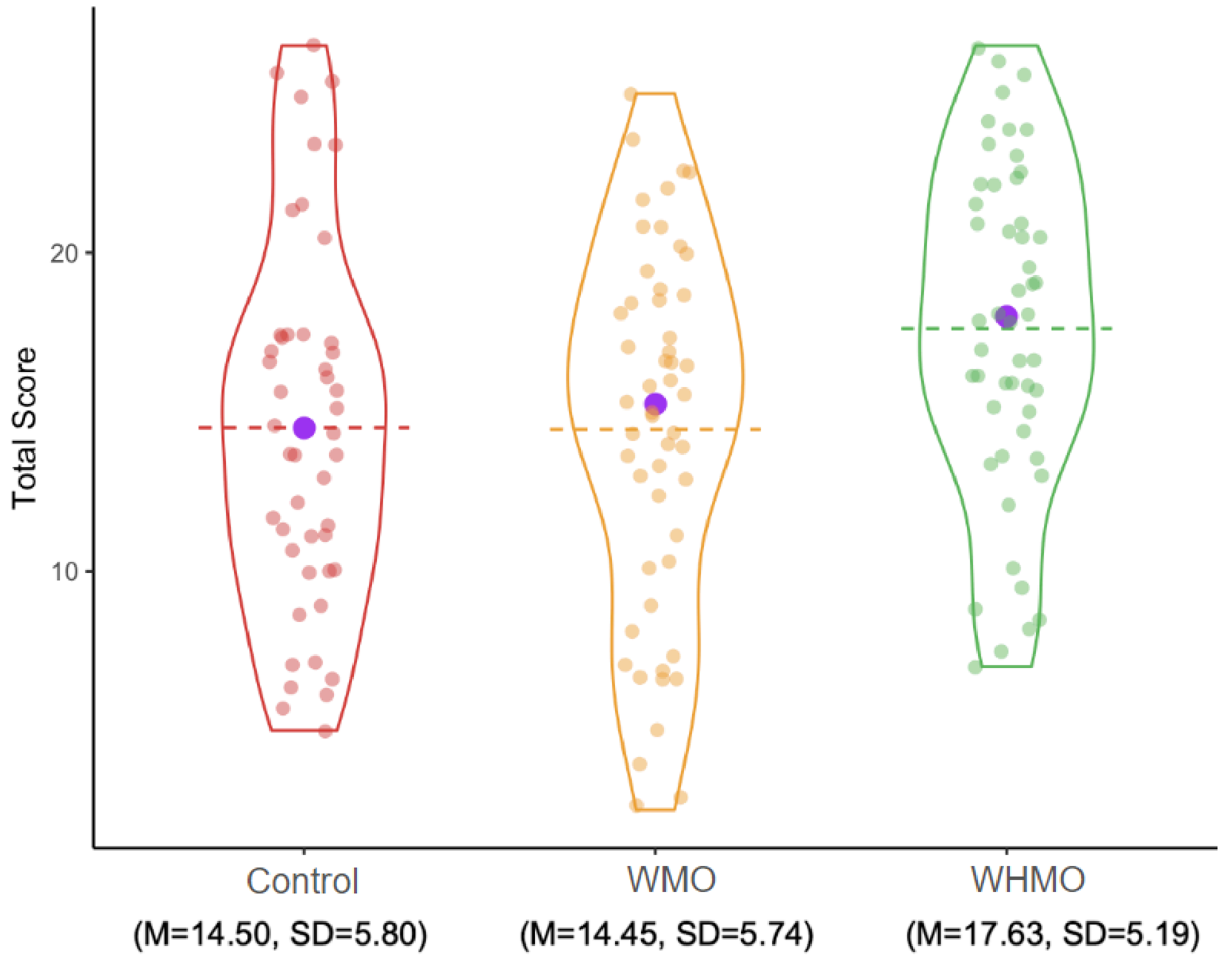
Results

Post-test Performance

Figure 5 shows the distribution of students' performance on the post-test questions by condition. Students' in the WHMO group scored higher on average than did students in either of the other two groups.

Figure 5

Violin Plots of Students' Posttest Performance by Condition



Note. Dashed lines are means; purple dots are medians.

An analysis of covariance (ANCOVA), which modeled post-test performance as a function of experimental condition while controlling for pre-test performance and study cohort (summer versus winter), revealed that the overall effect of condition significantly impacted post-test performance (Table 3).

Table 3

ANCOVA Results

Predictor	*df*	*F*	PRE	η^2_p	*95% CI for* η^2_p	*p*
Model (error reduced)	5	19.64*	0.41			<.001
Condition	2	6.38*	0.08	0.10	[0.02, 0.20]	.002
Pretest performance	1	79.92**	0.36	0.36	[0.24, 0.47]	<.001
Time (winter/summer)	2	0.07	0.00	0.00	[0.00, 0.00]	.934

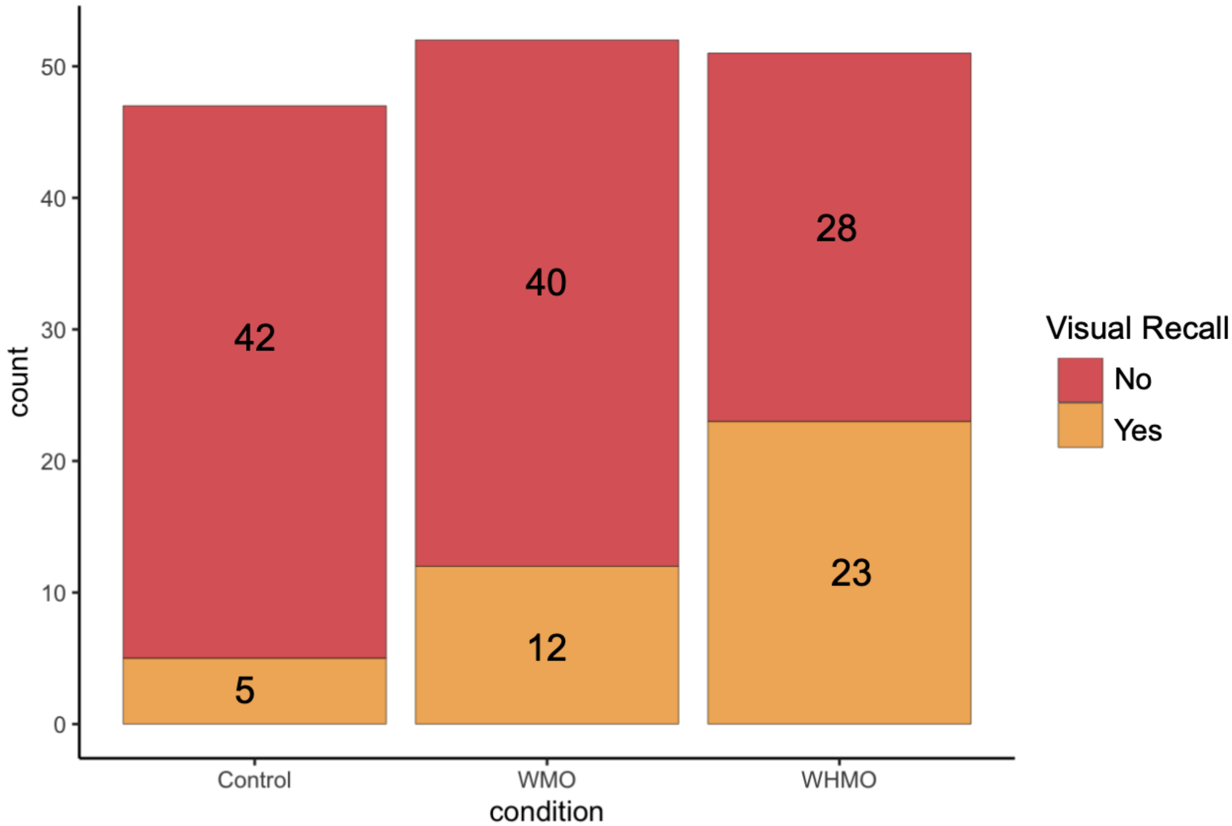
Post-hoc comparisons showed that students in the WHMO group (M = 17.63, SD = 5.19) outperformed both those in the WMO group (M = 14.45, SD= 5.74; $t(146)$ = 5.10, p_{adj}< .001) and those in the Control group (M = 14.50, SD= 5.80; $t(146)$ = 4.89, p_{adj}< .001). (Note: The error variance is pooled across all groups and then weighted to the groups being compared to offer a more robust error term; the p-values were adjusted for multiple comparisons using Bonferroni correction).

Did Participants' Visual Recall of the Video during Problem Solving Differ by Condition?

Figure 6 shows the number of participants who self-reported having a visual recall of the video during problem solving broken down by condition. A logistic regression showed that participants in the WMO group were 152% more likely to think back to visual components than those in the Control group (log odds = 0.92, odds ratio = 2.52, p = .109). Participants in the WHMO group were 590% more likely to think back to visual components than the Control group (log odds = 1.93, odds ratio = 6.90, p < .001), and 174% more likely than the WMO group (log odds= 1.01, odds ratio = 2.74 p = .020).

Figure 6

Participants' Thinking back to Visual Components by Condition

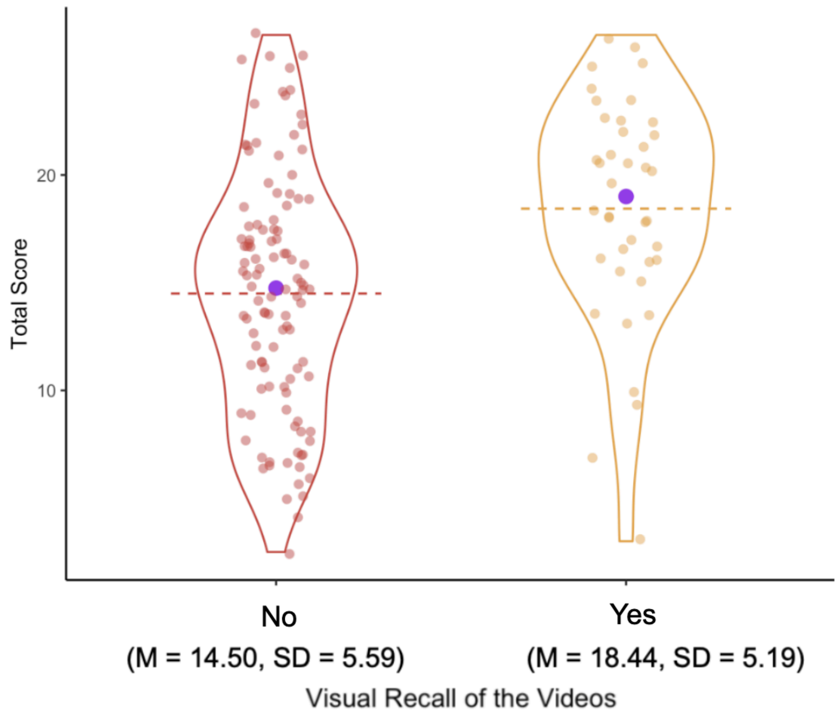


Did Visual Recall of the Video Predict Participants's Posttest Performance?

Figure 7 shows violin plots and descriptive statistics of participants' posttest performance separated by whether or not they reported a visual recall of the videos. An independent sample t-test showed that post-test performance of participants who mentioned a visual recall of the video was significantly higher than that of participants who did not recall visual components of the videos (t (148) = 3.89, $p < .001$).

Figure 7

Violin Plots of Posttest Score by Whether Participants Reported Having Visual Recall of the Videos



Note. Dashed lines are means; purple dots are medians.

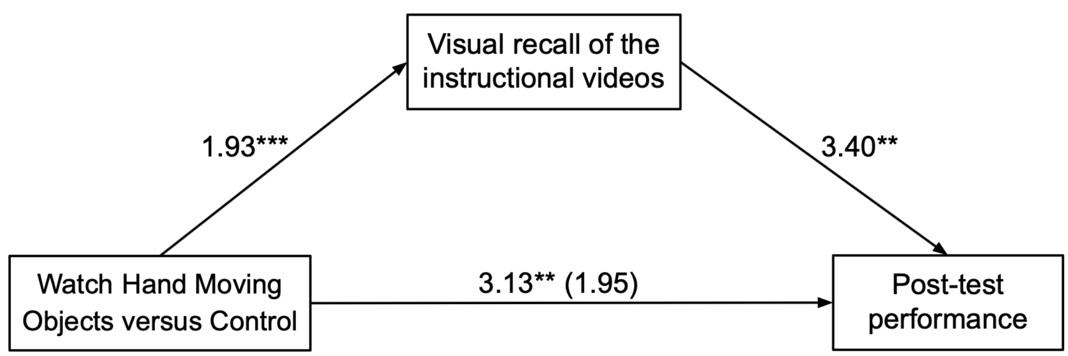
Did Visual Recall of the Video Mediate the Effect of Condition on Learning?

Because the mediator is binary, we used a causal mediation analysis to evaluate whether the effect of condition on posttest performance was significantly mediated by whether participants reported visual recall of the instructional videos during the posttest assessment. Because the predictor variable (i.e., condition) was multi-categorical with three levels, we fitted three mediation models.

The effect of the WHMO condition versus the Control condition on participants' posttest performance was significantly mediated by participants' self-reported visual recall (Figure 8; average causal mediation effect = 1.22, 95% CI with 5,000 non-parametric bootstrapping = [0.32, 2.12], p = .004). The effect of the WHMO condition versus the WMO condition on posttest performance was also significantly mediated by visual recall (Figure 9; indirect effect = 0.81, 95% CI with 5,000 non-parametric bootstrapping = [0.02, 1.67], *p* = .037). The effect of the WMO condition versus the Control condition on posttest through visual recall) was not significantly mediated (see Appendix C for complete results).

Figure 8

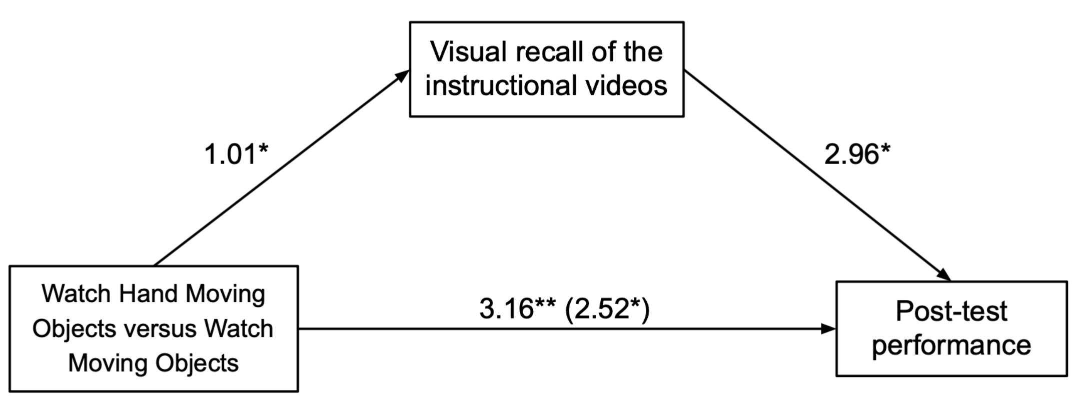
Diagram Showing Visual Recall of Videos as a Mediator of the Effect of Watching Hands Moving Objects (versus Control) on Posttest Performance



Note. The estimate for the a path is in the form of log odds.

Figure 9

Diagram Showing Visual Recall of Videos as a Mediator of the Effect of Watching Hands Moving Objects (versus Watching Moving Objects) on Posttest Performance



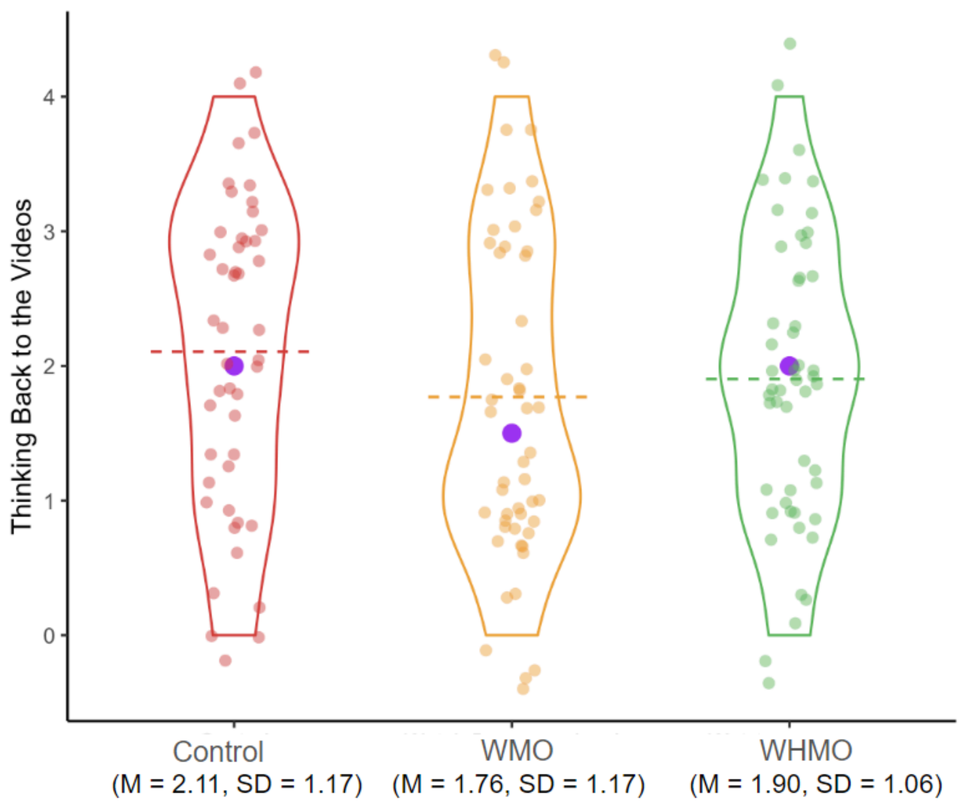
Note. The estimate for the a path is in the form of log odds.

Did Participants' General Recollection of the Video during the Posttest Differ by Condition?

In order to explore whether how often participants thought about the content in the videos differed by conditions, we performed a one-way ANOVA. The one-way ANOVA did not reveal a significant difference between the three conditions in terms of their general recollection of the video during the posttest (Figure 10; F (2,147) = .95, p = .389).

Figure 10

Participants' Thinking Back to the Video by Condition



Did Participants' General Recollection of the Video Predict their Posttest Performance?

There was no significant correlation between general recollection of video and posttest performance (t (148) = -0.08, p = .931).

Discussion

The current study explored whether sensorimotor engagement confers unique learning benefits over and above those provided by concreteness. Consistent with the embodied cognition hypothesis, the findings showed that the WHMO group outperformed the WMO group and the Control group. This distinction between visuospatial concreteness alone and the addition of sensorimotor engagement enables us to see that sensorimotor engagement confers a distinct advantage on top of visuospatial concreteness in promoting learning outcomes, whereas concreteness by itself, as implemented in the dynamic visualizations, is no more helpful than abstract demonstration. The mediation analyses further suggest that one potential mechanism for this effect might be that the hands-on demonstrations enabled students to more easily activate visual memories of the demonstrations, and then use these visual representations during problem-solving.

It is also interesting that students' performance after viewing the *concrete+dynamic* representations without hands (i.e. the WMO) was almost identical to their performance after viewing the live-coding control, which was both less concrete and less dynamic. In the realm of computer science education, live-coding demonstrations are thought to be better than simply showing students a large block of code on a slide. In live-coding, instructors can dynamically show and run each line of code one at a time, and model how coding behaviors unfold over time (Bennedsen & Caspersen, 2005). It is possible that because the live-coding videos dynamically showed the output of the code (i.e., printed out a changed dataset), this level of concreteness and dynamic quality was enough to produce a learning benefit. An additional dose of concreteness may not have added more value.

Previous research has produced mixed findings regarding the effectiveness of dynamic visualizations. Concerns have been raised regarding the cognitive load imposed by dynamic visualizations (Hegarty, 2004; Tversky et al., 2002). On one hand, the dynamic visualizations in the WMO and WHMO stimuli may have been quite similar in cognitive load because the content was highly similar. On the other hand, it may be that the embodiment involved in seeing the hands moving the objects made it easier for students to generate and sustain the visual and dynamic representations, and to use them in problem solving (de Koning & Tabbers, 2011; Zhang et al., 2022).

Only a few studies of embodied cognition have experimentally examined the intersection of sensorimotor engagement, visuospatial concreteness, and dynamic visualization, and each has focused on slightly different combinations of these features. Ours kept visuospatial concreteness and dynamic quality relatively constant while manipulating levels of sensorimotor engagement. This manipulation is similar to the comparison of performing versus observing gestures (e.g., Goldin-Meadow et al., 2012), but ours is the first to our knowledge to demonstrate the unique benefit of observing sensorimotor activities over concrete dynamic visualizations. Others have kept sensorimotor engagement and dynamic quality constant and manipulated levels of concreteness (e.g., Ping & Goldin-Meadow, 2008).

Beyond these few studies, other combinations of these three features of embodiment also merit attention. For example, we need studies that keep sensorimotor engagement and concreteness constant while manipulating the dynamic quality of the representations. Would participants benefit more from a video of an instructor's hand drawing a diagram than from a video in which the instructor's hand simply pointed to parts of an already finished drawing? We have begun to explore this possibility by teaching students about the normal distribution in statistics – a diagram that instructors commonly draw and point to as they teach students (Zhang et al, in press). Future research could also investigate *how much* of these features are needed in order to benefit learning. In the current study we included only the hands in our sensorimotor engagement condition; would including the entire body result in different benefits?

The disentanglement enabled by the current study highlights the need for a more nuanced understanding of what we mean when we say "embodied learning experiences." The three features we have proposed are only a starting point to finding the best ways to characterize a construct as complex as embodied learning. Whether embodiment should be categorized in types or considered on a continuum, ranging from purely abstract representations to somewhat embodied ones (e.g., observing actions) to strongly embodied ones (e.g., performing actions), varies depending on the theory and remains a subject of inquiry (e.g. Johnson-Glenberg and Megowan-Romanowicz, 2017). Additionally, it remains unclear whether the learning benefits of embodiment increase linearly with the levels of embodiment.

When there are learning benefits of embodiment, what are the mechanisms underlying the effect? The robustness and content of mental representations may be one potential mechanism for the benefits of embodied learning. This study not only narrowed in on the causal relationship between sensorimotor engagement and learning benefits but also revealed that sensorimotor engagement prompted learners to activate and use visuospatial mental representations, ultimately resulting in enhanced learning outcomes. Recognizing that sensorimotor engagement provides a distinct benefit has the potential to reshape instructional practices and curriculum development, moving beyond a narrow emphasis on visuospatial concreteness and expanding to incorporate deliberate engagement of the body and physical experiences in the world.

Constraints on Generality

The participants in this study are students from a highly selective public university learning coding and statistics. Compared to the general student population in the United States, they are students who are high-achieving and have relatively high content knowledge. Future studies might want to implement a similar design on students from more diverse educational backgrounds. Further research is also needed to explore the implications of sensorimotor engagement and perceptual concreteness in domains beyond coding and statistics.

In conclusion, this study highlights that students exposed to hands-on representations exhibited superior learning outcomes because of the unique contribution of sensorimotor engagement beyond perceptual concreteness and dynamic quality. By exposing the role of mental representations in embodied benefits to learning, this study sheds light on the processes underlying embodied learning. Finally, the practical implications for teaching are noteworthy. As educators face daily decisions regarding the integration of different types of representations and activities into their lessons, this study advocates for the inclusion of bodily movements, even in lessons that are already perceptually concrete.

Appendix A

1. In your own words, explain what the shuffle() function does
2. In your own words, explain when you would use the shuffle() function
3. Which process do you think will create a more random result? Shuffling once or ten times? Explain your answer
4. Given a specific dataset, would the number of observations in the condition variable (either experimental or control) increase, decrease, stay the same or we cannot know until after we see the shuffled result after the condition column is shuffled
5. Suppose you roll a dice four times, which is more likely to occur and why: see the numbers 6, 6, 6, 6 in order, see the numbers 1, 2, 3, 4 in order or see the numbers 3, 4, 1, 6 in order.

Appendix B

Now, you will answer some questions based on the videos you have watched.

The laptop_data dataset contains data from an experiment on the effect of laptops on student learning. Undergraduate students were randomly assigned to one of two conditions: view or no-view. In the view condition, students attended a 40 minute lecture and were allowed to keep their laptops open. In the no-view condition, students attended the same lecture, but were asked to keep their laptops closed. At the end of the lecture, students took a test on the lecture content and rated how distracted they felt during class.

There are three variables in this dataset:

- condition: the condition students were randomly assigned to, either view or no-view
- total: the percentage of questions students answered correctly on the post-lesson assessment
- distracted: students' self-reported rating of how distracted they were in class.

1. What would you expect to happen to the value of **condition** for row 1 if we ran the code below?

```
laptop_data$condition <- shuffle(laptop_data$condition)
```

____

2. What would you expect to happen to the value of condition for row 1 if we instead ran the code below?

   laptop_data$total <- shuffle(laptop_data$total)

---

3. We ran this code to create a table that shows the number of observations in each condition.

```
tally(~ condition, data = laptop_data)
```

```
condition
no-view    view
     19      19
```

Now, imagine we run this code:

```
laptop_data$condition <- shuffle(laptop_data$condition)
tally(~ condition, data = laptop_data)
```

What would happen to the number of observations in the view condition?

The number of observations would increase

a. The number of observations would stay the same
b. The number of observations would decrease
c. The number of observations would increase, decrease, or stay the same, but it's impossible to tell which

4. Explain your answer to the previous question

____

We used the code below to create a faceted histogram showing the distribution of total in each condition. The vertical lines represent mean total scores for the two conditions. Again, you can see that the participants in the no-view group scored higher, on average, than participants in the view group.

```
stats <- favstats(total ~ condition, data = laptop_data)
```

```
 gf_dhistogram(~ total, data = laptop_data) %>%
   gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
   gf_facet_grid(condition ~ .)
```

5. Sometimes groups differ just because of randomness. Do you think the group difference in the histogram above could be due to randomness?
   a. Yes, it must be due to randomness
   b. No, it cannot be due to randomness
   c. Maybe, need to further investigate
6. Explain your answer to the previous question

____

7. If you wanted to investigate whether this difference could be due to randomness, what would you do? Please be as specific as possible in your response.

____

8. Alex thinks she only needs to shuffle once to see if the difference between conditions on total could be due to randomness by comparing the shuffled result with the original data. Mary thinks she needs to shuffle more than once to be able to see if the difference could be due to randomness. Do you agree with Alex or Mary? Explain your answer.

____

Take a look at each line of code below. For each line, **explain 1) what the code is doing** and **2) why someone would write that code**.

```
laptop_data$condition.shuffle <- shuffle(laptop_data$condition)
```

9. What is this line of code doing?

____

10. Why would someone write this line of code?

____

```
laptop_data$total.shuffle<- shuffle(laptop_data$total)
```

11. What is this line of code doing?

____

12. Why would someone write this line of code?

13. Look at the two examples of codes below. Example 1 and Example 2 each produces a faceted histogram. In what ways would the two faceted histograms be similar?

Example 1:

```
gf_dhistogram(~ distracted , data = laptop_data) %>%
  gf_facet_grid(shuffle(condition) ~ .)
```

Example 2:

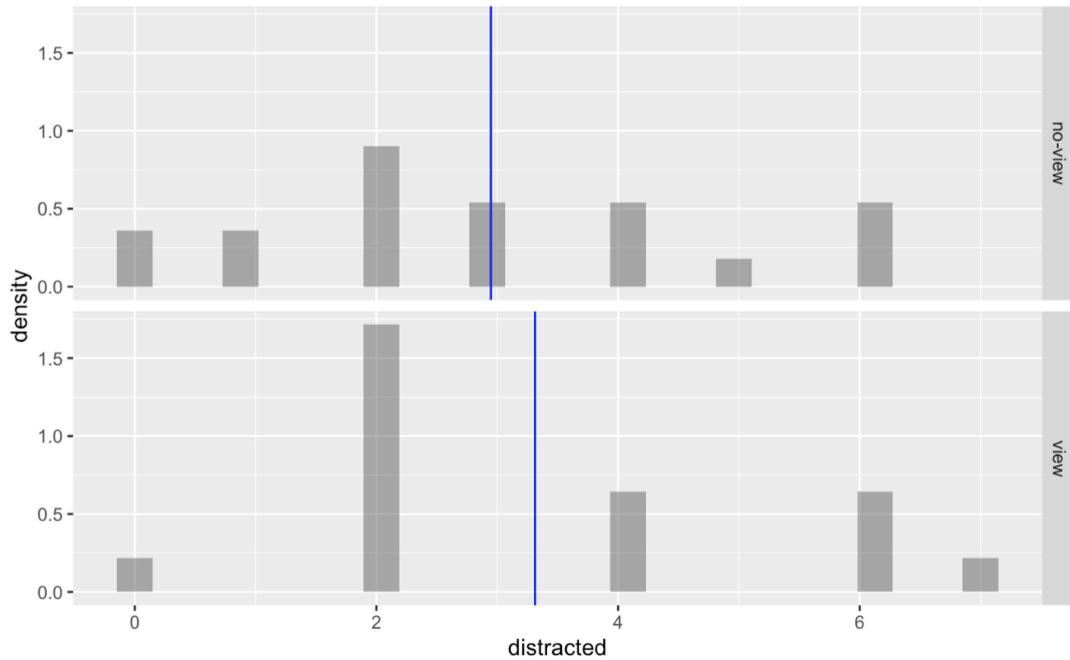
```
gf_dhistogram(~ shuffle(distracted) , data = laptop_data) %>%
  gf_facet_grid(shuffle(condition) ~ .)
```

14. Would one histogram be more random than the other one? If yes, which one is more random and why; if not, why not.
15. Would the two histograms look exactly the same or different? Explain your answer.

We ran this code to create the graph below. We added a line in each condition to represent the mean of **distracted** of that **condition**. Notice that the average **distracted** rating in the **no-view condition** is lower than the average **distracted** rating in the **view condition**.

```
stats <- favstats(distracted ~ condition, data = laptop_data)
gf_dhistogram(~ distracted, data = laptop_data) %>%
```

```
gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
gf_facet_grid(condition ~ .)
```



16. Sometimes groups differ just because of randomness. Do you think the group difference in the histogram above could be due to randomness?
    a. Yes, it must be due to randomness
    b. No, it cannot be due to randomness
    c. Maybe, we need to further investigate
17. Explain your answer to the previous question

____

18. If you ran the code in the previous question again, do you think it would produce the same output?
    a. Yes
    b. No

c. It's possible, but not likely

19. Explain your answer to the previous question

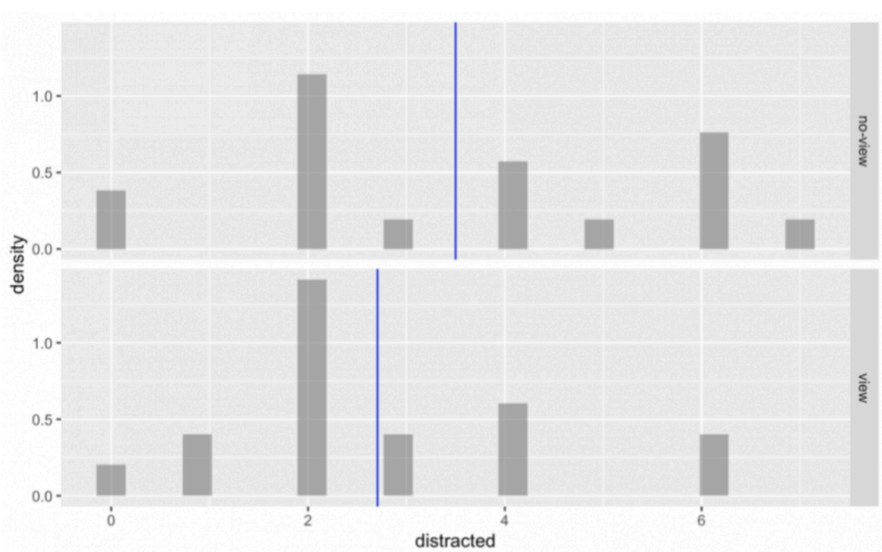
____

We revised the code from the previous question to create the graph below. We added a line to represent the mean of **distracted** for each **condition**. Notice that the average **distracted** rating in the **no-view condition** is higher than the average **distracted** rating in the **view condition**.

20. What caused the difference in the means represented in the graphs below?

```
laptop_data$condition.shuffle <- shuffle(laptop_data$condition)
stats <- favstats(distracted ~ condition.shuffle, data = laptop_data)

gf_dhistogram(~distracted, data = laptop_data) %>%
  gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
  gf_facet_grid(condition.shuffle ~ .)
```



____

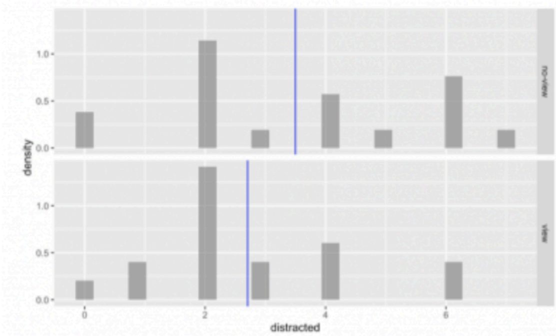
21. Sometimes groups differ just because of randomness. Do you think the group difference in the histogram above could be due to randomness?
    a. Yes, it must be due to randomness
    b. No, it cannot be due to randomness
    c. Maybe, need to further investigate
22. Explain your answer to the previous question

    ____

23. If you ran the code in the previous question again, do you think it would produce the same output?
    a. Yes
    b. No
    c. It’s possible, but not likely
24. Explain your answer to the previous question

    ____

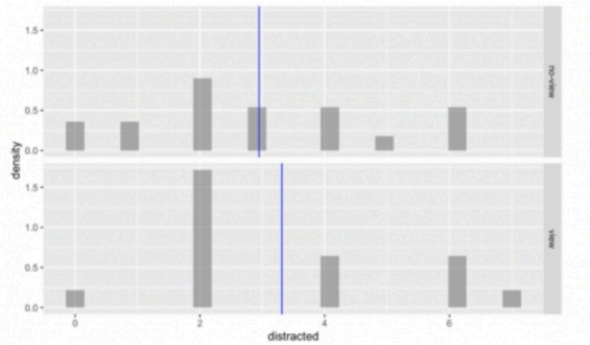
Look at the two faceted histograms below, along with the code that produced each (the code might be a bit hard to read, feel free to zoom in to get a better read):




```
laptop_data$condition.shuffle <- shuffle(laptop_data$condition)

stats <- favstats(distracted ~ condition.shuffle, data = laptop_data)

gf_dhistogram(~distracted, data = laptop_data) %>%
gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
gf_facet_grid(condition.shuffle ~ .)
```




```
stats <- favstats(distracted ~ condition, data = laptop_data)

gf_dhistogram(~ distracted, data = laptop_data) %>%
gf_vline(xintercept = ~mean, data = stats, color = "blue") %>%
gf_facet_grid(condition ~ .)
```

25. Why do the two faceted histograms look different?

____

26. Based on what you've learned from these two histograms, do you think being able to view or not view a laptop during class (condition) affects students' self-reported rating of how distracted they were in class (as measured by distracted score on a post-lesson assessment)? Why or why not?

____

Imagine we run the code below:

laptop_data$distracted.shuffle <- shuffle(laptop_data$distracted)

mean(laptop_data$distracted.shuffle)

mean(laptop_data$distracted)

27. How would the mean of distracted.shuffle compare to the mean of distracted?

a. The mean of distracted.shuffle would be larger
b. The mean of distracted.shuffle would be smaller
c. The two means would be the same
d. It's impossible to tell

28. Explain your answer to the previous question

____

29. What will the distribution of the variable, distracted.shuffle, look like compared to the distribution of the variable, distracted?
    a. Wider
    b. Narrower
    c. The same
    d. Not sure. It will vary randomly

30. Explain your answer to the previous question

____

Imagine now we have a new variable, gender, so that we have four variables in the dataset:

gender: the gender students self-identify with

condition: the condition students were randomly assigned to, either view or no-view

total: the percentage of questions students answered correctly in their final exam

distracted: students' self-reported rating of how distracted they were in class.

31. If we now shuffle the column of gender, what would happen to the relationship between condition and total? Explain your answer.

____

32. What do you think the purpose of the shuffle() function is?

____

33. In your own words, explain when you would use the shuffle() function.

____

Appendix C

The indirect effect of watching moving objects without hands versus watching live-coding on participants' posttest performance through participants' self-reported visual recall was not statistically significant (indirect effect = 0.29, 95% CI with 5,000 non-parametric bootstrapping = [-0.20, 1.34], $p = .16$).

References

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and cognitive processes*, *15*(6), 593-613.

Bennedsen, J., & Caspersen, M. E. (2005, February). Revealing the programming process. In *Proceedings of the 36th SIGCSE technical symposium on Computer science education* (pp. 186-190).

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.

Brooks, N. B., Barner, D., Frank, M., & Goldin‐Meadow, S. (2018). The role of gesture in supporting mental representations: The case of mental abacus arithmetic. *Cognitive science*, *42*(2), 554-575.

Castro-Alonso, J. C., Ayres, P., & Sweller, J. (2019). Instructional visualizations, cognitive load theory, and visuospatial processing. *Visuospatial processing for education in health and natural sciences*, 111-143.

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & De Rosario, H. (2017). pwr: Basic functions for power analysis (Version: 1.3-0) [Software]. https://cran.r-project.org/we b/packages/pwr/index.html

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, *137*(4), 706.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences New York.* Academic.

Cook, S. W., Friedman, H. S., Duggan, K. A., Cui, J., & Popescu, V. (2016). Hand gesture and mathematics learning: lessons from an avatar. *Cognitive Science*, *41*(2), 518–535. https://doi.org/10.1111/cogs.12344.

de Koning, B. B., & Tabbers, H. K. (2011). Facilitating understanding of movements in dynamic visualizations: An embodied perspective. *Educational Psychology Review*, *23*, 501-521.

Fiorella, L., & Mayer, R. E. (2016). Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology*, *108*(4), 528.

Fyfe, E. R., McNeil, N. M., & Borjas, S. (2015). Benefits of "concreteness fading" for children's mathematics understanding. *Learning and Instruction*, *35*, 104-120.

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*, 9-25.

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, 5(6), 664–674. https://doi.org/10.1177/1745691610388764.

Goldin‐Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental science*, *15*(6), 876-884.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological science*, *12*(6), 516-522.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, *65*(2-3), 231-262.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in cognitive sciences*, *8*(6), 280-285.

Johnson-Glenberg, M. C., & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive research: principles and implications*, *2*(1), 1-28.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of educational psychology*, *106*(1), 86.

Johnson-Glenberg, M. C., Megowan-Romanowicz, C., Birchfield, D. A., & Savio-Ramos, C. (2016). Effects of embodied learning and digital platform on the retention of physics content: Centripetal force. *Frontiers in psychology*, *7*, 1819.

Kazak, A. E. (2018). Editorial: Journal article reporting standards.*American Psychologist, 73*(1), 1–2. https://doi.org/10.1037/amp0000263

Montessori, M. (1917). *The advanced Montessori method* (Vol. 1). Frederick A. Stokes Company.

Nathan, M. J., & Alibali, M. W. (2011). How gesture use enables intersubjectivity in the classroom. *Integrating gestures*.

Novack, M., & Goldin-Meadow, S. (2015). Learning from gesture: How our hands change our minds. *Educational psychology review*, *27*, 405-412.

Novack, M. A., Wakefield, E. M., & Goldin-Meadow, S. (2016). What makes a movement a gesture?. *Cognition*, *146*, 339-348.

Piaget, J. (1970). *Genetic epistemology*. Columbia University Press.

Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology*, *44*(5), 1277–1287. https://doi.org/10.1037/0012-1649.44.5.1277.

Pouw, W. T. J. L., van Gog, T., Zwaan, R. A., & Paas, F. (2016). Augmenting instructional animations with a body analogy to help children learn about physical systems. *Frontiers in Psychology*, *7*, 860.

Pruim, R. J., Kaplan, D. T., & Horton, N. J. (2017). The mosaic package: helping students to'think with data'using R. *R Journal*, *9*(1), 77.

R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Rimé, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, *8*, 311-325.

Schachner, A., & Carey, S. (2013). Reasoning about 'irrational'actions: When intentional movements cannot be explained, the movements themselves are seen as the goal. *Cognition*, *129*(2), 309-327.

Shapiro, L., & Stolz, S. A. (2019). Embodied cognition and its significance for education. *Theory and Research in Education*, *17*(1), 19-39.

Tran, C., Smith, B., & Buschkuehl, M. (2017). Support of mathematical thinking through embodied cognition: Nondigital and digital approaches. *Cognitive Research: Principles and Implications*, *2*, 1-18.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate?. *International journal of human-computer studies*, *57*(4), 247-262.

Uttal, D. H., Liu, L. L., & DeLoache, J. S. (2006). Concreteness and symbolic development. *Child psychology: A handbook of contemporary*,(2nd), 167-184.

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial?. *Journal of memory and language*, *50*(4), 395-407.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Zhang, I. Y., Givvin, K. B., Sipple, J. M., Son, J. Y., & Stigler, J. W. (2021). Instructed hand movements affect students' learning of an abstract concept from video. *Cognitive science*, *45*(2), e12940.

Zhang, I. Y., Tucker, M. C., & Stigler, J. W. (2022). Watching a hands-on activity improves students' understanding of randomness. *Computers & Education*, *186*, 104545.

Chapter 3

The Role of Prior Knowledge in Effects of Embodied Pedagogies on Learning

Icy Zhang, Alice Xu, Ji Y. Son, Craig Enders, and James W. Stigler

University of California, Los Angeles

Abstract

In this study, we implemented and compared two types of embodied pedagogy: one in which students actively participated in hands-on activities, and another in which they observed other students doing the activities. The activities were implemented as the lab component of a 10-week college-level introductory statistics course in which students were taught to use R for data analysis. 227 college students participated in the study. Half were randomly assigned to perform a series of hands-on activities and half to observe a partner performing the same activities. We hypothesized that students with less prior knowledge would benefit more from active participation, while those with more prior knowledge would gain more from observing their peers, a hypothesis we call the Performing First Hypothesis. As predicted, an analysis of students' exam scores showed a significant interaction between students' prior knowledge (as measured by a self-rating) and type of pedagogy in the hypothesized direction. Interestingly, only self-rated prior knowledge—not factors like math anxiety or previous math performance—significantly moderated the effectiveness of the embodied pedagogies. Multilevel analysis of weekly pretests and delayed posttests confirmed these findings, supporting the Performing First Hypothesis and suggesting new directions for research on the mechanisms and applications of embodied pedagogies in complex learning environments.

The Role of Prior Knowledge in Effects of Embodied Pedagogies on Learning

The idea that bodily experiences enhance learning is not new. Embodied learning pedagogies, the use of the body through learners' enactment, observation, and even mental simulation, have emerged as a promising approach to fostering transferable knowledge (Wilson, 2002). Past research has observed an increase in learning when learners directly act on external manipulatives or representations (Sommerville & Woodward, 2010; Wilson, 2002), represent ideas in gesture or directed actions without exerting any change to the external environment (Congdon & Goldin-Meadow, 2021; Zhang et al., 2021), or simply watch someone else produce gestures or object manipulations (Cook et al., 2024; Zhang et al., 2022). However, our understanding of how embodied pedagogies enhance learning in complex domains like mathematics or statistics remains nascent. Although these methods have been well-explored in lab settings, mastering complex real-world domains requires more extensive interaction with a broad array of interconnected concepts over longer periods of time than is typically offered in laboratory studies (Fries et al., 2021).

This paper aims to address this gap in our current understanding by posing two key questions. The first question focuses on the generalizability of embodied pedagogies beyond the learning of single concepts. Most of the evidence for embodied pedagogies comes from laboratory studies. Although embodied pedagogies have been shown to facilitate the learning of a single concept in the lab, how applicable and impactful are they in a course setting where students are expected to learn a multitude of interrelated concepts over time? Studying embodied pedagogies in a longitudinal environment allows us to more closely examine learners' knowledge development in ways that we could not in laboratory settings.

The second question concerns individual differences among learners, particularly the diversity of their prior knowledge. In lab studies, prior knowledge is usually treated as a covariate and only a few studies have considered how prior knowledge could interact with embodied pedagogies (e.g. Congdon et al., 2018; Cook et al., 2024; Guarino & Wakefield, 2020; Zacharia et al., 2012). Even laboratory experiments that do consider prior knowledge may not be a good model for investigating how learners develop a deep understanding of abstract concepts in STEM domains. Although research has shown that when relevant prior knowledge is activated during learning, it contributes to the development of later concepts (Brod, 2021; Thompson & Zamboanga, 2003; Simonsmeier et al., 2022), the contribution of prior knowledge might be different in a longitudinal setting than in a single lab experiment. The critical point is that when learners come into the class for an embodied learning experience, their understanding of the concepts differs. No study to our knowledge has directly investigated how prior knowledge might moderate the effect of embodied pedagogies in complex domains over a significant period of time. How do interventions with different levels of embodiment impact learners with different levels of prior knowledge and experiences in the domain?

A focus on this question has the potential to serve students with a wide range of prior educational experiences. In particular, we identify the subject of statistics and data science. In this increasingly data driven society, it is important for students to use programming technology to explore messy data landscapes and make data-informed decisions, but because concepts in this domain are intrinsically abstract, novices frequently fail to develop interconnected robust mental representations of these concepts (Lau & Yuen, 2008; West & Ross, 2002). Could embodied learning be an effective strategy to serve students with low prior knowledge of statistics and programming? Beyond practical insights, this line of investigation is also important from a theoretical standpoint: do different types of embodied pedagogies differ in their effect on people with varying levels of prior experience? If so, that would shed light on the potential mechanisms that may underlie the benefits of embodied pedagogies.

To begin answering these two questions, we build up a theoretical framework based on current literature looking for clues about how embodied pedagogy might impact long-term learning in an academic domain with a diversity of learners. Then we detail the results of an in-class intervention experiment that exposed students to embodied learning activities for nine weeks.

Theoretical Framework

Theories of embodied cognition assume that concepts are grounded, given meaning, and intertwined with action and perception (Barsalou, 1999, 2008; Borghi & Pecher, 2011; Clark, 2008; Golonka & Wilson, 2012; Piaget, 1983). According to these theories, physical actions observed and performed during learning influence our internal mental representations as well as subsequent information processing, including problem-solving, reasoning, and retrieval (Barsalou, 2008; Fu & Franz, 2014). Given this, it is essential that classroom teachers carefully consider how to incorporate embodied actions, as overlooking this key component could significantly diminish the effectiveness of the learning experience (Abrahamson et al., 2020; Sullivan, 2018).

STEM education has emerged as a particularly promising area for investigating embodied instructional approaches. Despite the inherent abstractness of many STEM topics, characterized by formal notation (e.g., symbols used in mathematical equations or to reference chemical elements) and not easily perceptible concepts, the efficacy of various embodied learning interventions in STEM domains has been validated through controlled laboratory experiments

(Goldin-Meadow & Wagner, 2005; Johnson-Glenberg & Megowan-Romanowicz, 2017; Zhang et al., 2021, 2022).

One Potential Mechanism Underlying the Effectiveness of Embodied Pedagogies

One potential explanation for these effects is that involving learners' bodies during instruction, whether through performing or observing actions, generates concrete sensorimotor experiences that imbue abstract STEM symbols with meaning. Both theory and empirical research have suggested the importance of "meaning" for embodied pedagogies to be effective. For example, the perceptual symbols system theory posits that when people first have a meaningful perceptual experience, it activates a combination of neurons that are stored mentally as multimodal frames of perceptual symbols. Later, observing or performing bodily actions activates these multimodal frames (in terms of neural connections) to create simulations of perceptions and actions that have happened in the past (Barsalou, 1999). The underlying assumption for the sensorimotor simulation relies on having a meaningful experience, i.e., a simulator, in the past that can be activated, without which the embodied experience loses meaning.

This idea is supported by infant research showing that what infants see when they observe an action (e.g., a hand reaching for an object) depends on whether they have experience performing such an action themselves. In a classic study, Sommerville and colleagues (2005) had infants who could not yet grasp an object wear velcro mittens, which, because of their stickiness, enabled infants to experience what it would be like to reach out and grasp an object. These infants, having now experienced the action themselves, found observing an experimenter perform a similar action more interesting than did infants without prior experience, as the action now held meaning for them (Gerson & Woodward, 2014).

We would argue that this same idea applies when students are learning abstract and difficult concepts in educational settings. Our own experience as well as research by others indicates that the same embodied pedagogy may be experienced as meaningful by some learners but not by others. Further, only when the embodied experiences are meaningful to learners do such experiences facilitate learning, presumably by helping them establish a robust mental representation of the concept (Authors et al., under review; Congdon et al., 2018; Cook et al., 2024). The critical question is what does it take for different types of embodied pedagogies to be meaningful for different learners? In order to approach this question, we need to first unpack the concept of “embodiment” or “embodied pedagogies” and then ask whether different embodied pedagogies impact different learners equally.

Distinguishing between Performing and Observing Bodily Actions

By now, a large number of studies have found that embodied pedagogies are better than non-embodied pedagogies (e.g., for reviews, see Nathan, 2021 and Novack & Goldin-Meadow, 2015). However, what counts as "embodied" varies across studies. In some studies, learners themselves perform bodily movements (e.g., Johnson-Glenberg et al., 2014), gestures (e.g., Broaders et al., 2007), object manipulations (Rosenbaum et al., 2012), and drawing (Zhang et al., 2024), whereas in other studies, learners merely observe actions performed by an instructor or another person (Cook et al., 2024; Goldin-Meadow et al., 2012; Zhang et al., 2022).

We argue that the distinction between performing and observing bodily actions is an important one, supported both by theory and research. According to the previously discussed Barsalou’s perceptual symbol systems theory (1999, 2008), there is a difference between performing and observing an action: performing is more likely to be meaningful to novice learners than observing.Observing will only be meaningful to the learner if there are existing multimodal frames to be activated and used as simulators that reactivate the same neural connections and thus simulate the original experience without bodily actions (Pezzulo & Calvi, 2011).

Neurological research on a class of motor neurons called mirror neurons provides support for Barsalou's theory (Rizzolatti & Craighero, 2004). Mirror neurons discharge when a person acts, but whether they will also discharge when a person observes an action depends on whether the observer understands the goal of the action (Calvo-Merino, 2013; Gazzola et al., 2007). In other words, they must find the action to be meaningful. If they don't understand the goal, simply observing may not be sufficient for them to develop an embodied representation (Sommerville et al., 2005).

Research on learning has shown that at least in some cases performing is more advantageous than observing. For example, Goldin-Meadow and colleagues (2012) directly compared the effects of performing versus observing gestures on a learning outcome. In this study, six-year-olds were asked on a mental rotation task to judge whether two shapes at different angles of rotation were the same or different. Children who were instructed to perform gestures that mimicked the rotation of the figures performed better on the task than children who simply observed someone else performing similar gestures on a video clip.

Research on the Role of Prior Knowledge in Embodied Pedagogies

Zacharia and colleagues (2012) demonstrated that learners with low prior knowledge benefited more from physically manipulating objects than from virtually manipulating objects on a computer screen. Kindergarteners who had a correct understanding of a balance scale (i.e. *high* prior knowledge) and those who did not (i.e., *low* prior knowledge) were randomly assigned to either physically interact with a balance scale and objects of different weights or virtually interact with those same objects on a screen. Arguably, physical manipulation is more perceptually available than virtual manipulation because it can be more directly experienced.

Zacharia and colleagues (2012) found that for participants who had *high* prior knowledge of a balance scale (i.e. a correct understanding), performance in using the balance scale to measure and differentiate objects of different mass did not differ based on whether they engaged in physical or virtual manipulation. However, participants with low prior knowledge demonstrated a different pattern of results: those who performed the physical manipulation outperformed those who performed the virtual manipulation. Notably, the low-prior-knowledge participants who did the virtual manipulation did not improve from the pre- to post-test, whereas the low-prior-knowledge participants who went through the physical manipulation improved the most.

Similarly, learners' prior knowledge mattered when researchers compared a more concrete gesture instruction (pretending to pick up and manipulate the addends in a math problem) with an abstract gesture instruction (a v-point gesture for mathematical equivalence followed by a single-point for the blank to fill in the answer). Children who learned with the abstract gesture instruction were more likely to transfer their knowledge to solve new problems (Novack et al., 2014). However, a secondary analysis of the data suggests an interaction between prior knowledge and type of instruction (Congdon & Goldin-Meadow, 2021). Children with low pretest scores learned the least from abstract gesture instruction compared with high-pretest children in the abstract gesture group and both low- and high-pretest children in the concrete-gesture group. The findings suggest that more perceptually available concrete gesture instruction might create a more meaningful mental representation for low-prior knowledge learners while the abstract gesture instruction might help high-pretest children to transfer their knowledge.

Whereas these two studies suggest that low prior knowledge learners benefit more from more embodied pedagogies, there is also some evidence suggesting that learners may need some prior knowledge in a domain just to benefit from embodied pedagogies. For example, Wakefield and James (2015) randomly assigned children to learn the concept of a palindrome with either speech-only instruction, speech+gesture match instruction or speech+gesture mismatch instruction. They found that only children with high phonological ability, but not children with low phonological ability, benefited more from training that required them to perform gestures (i.e. speech+gesture match and speech+gesture mismatch) than from training that included speech-only instruction. A later paper by Guarino and Wakefield (2020) interpreted this finding in terms of children's developing knowledge in the domain. They posited that there is a developmental point for children to benefit from gestures and argued that very young children may lack necessary existing knowledge in the domain to understand how gesture can index speech. On the other hand, older children with sufficient existing knowledge may not not need gestures to understand.There might be a "sweet spot" where children possess the foundational knowledge and cognitive ability for gestures to benefit word learning.

However, learning abstract concepts in complex domains is a very different context than the one studied by Wakefield and James. Especially if the learners are students enrolled in a college-level class, as the course progresses, they should all develop some foundational knowledge and all possess a higher level of cognitive ability than that of young children. It is still unclear what the effect of embodied pedagogies would be on adults learning abstract concepts in complex domains with different levels of prior knowledge, who are developmentally different from children.

Past research suggests prior knowledge is also a robust predictor of learning in adults (Chakraborty & Esposito, 2024); see Tobias, 1994, for review) and might moderate the effect of learning interventions. One important study by Grotelueschen (1979) found that adults with low prior knowledge in the domain benefited more from an instructional sequence that went from concrete to abstract whereas adults with high prior knowledge benefited more from materials that were abstrac throughout. But certainly more research is needed to elucidate the role of prior knowledge in embodied learning in adults.

The Performing First Hypothesis

Building on the distinction between performing and observing actions and the importance of meaning previously established in this paper, we propose a hypothesis that we will refer to as the Performing First Hypothesis. This hypothesis is that performing bodily actions will benefit learners with no or low prior knowledge in a domain more than learners with high prior knowledge, whereas observing bodily actions will benefit learners with high prior knowledge more than learners with no or low prior knowledge. Whereas most research on educational interventions finds a positive correlation between students' prior knowledge of the concepts and their performance after the intervention, performing embodied activities might attenuate or even eliminate the correlation between prior knowledge and performance after the intervention.

Although the specific interaction between prior knowledge and performing vs. observing has not been tested, some empirical support for this hypothesis may be found in research comparing other types of embodied pedagogies. For example, Congdon et al. (2018) compared gesture and object manipulation to see whether learners with low prior knowledge might understand one but not the other. Because object manipulation enables learners to directly interact with external objects, it is more perceptible and potentially more comprehensible by novice learners. Congdon and colleagues found that when students learning linear units of measure were randomly assigned to one of the two embodied pedagogies, those with low prior knowledge benefited from object manipulation but not from gesture. In contrast, children with higher prior knowledge benefited equally from both types of embodied pedagogies.

This difference suggests that whereas *object manipulation* can be more directly experienced and is thus more meaningful to novice learners, *gesture* might only activate embodied knowledge for learners who have the prior knowledge required to make meaningful connections between the gesture and the targeted abstract concepts. This finding provides rudimentary support for the Performing First hypothesis. Higher prior knowledge makes less perceptible embodied pedagogies meaningful to learners, whereas learners with low prior knowledge need more perceptible experiences, such as physically manipulating the objects, to understand the meaning behind the actions.

Both of the embodied pedagogies included in the Congdon et al. study involved physical performance. We do not know the effect that merely *observing* embodied pedagogies would produce for learners with different levels of prior knowledge. Findings from this study also do not speak to the other half of the Performing First hypothesis, that is, whether learners with high prior knowledge will benefit more from observing embodied pedagogies or whether more concrete and perceptible embodied pedagogies are always better for learning. At least some evidence reviewed above suggests that more concrete pedagogies might diminish learners' ability to transfer knowledge to solve new problems in novel contexts (Novack et al., 2014).

In addition, these pieces of evidence all come from highly controlled lab experiments and from children who differ both developmentally and cognitively from college students. Moreover, past research only measured participants' knowledge right before and right after the intervention. Because developing expertise in STEM domains is a lengthy process, a single intervention may not accurately reveal how prior knowledge and types of embodied intervention might interact over time.

The Current Study

In the current study, we developed a supplementary lab curriculum employing embodied pedagogies, and then implemented the curriculum over a 10-week term as part of a college-level introductory course in statistics and data science. The course was taught using a CourseKata interactive online textbook, available for preview at https://coursekata.org. The book takes a modeling approach to statistics, and teaches students to analyze data using R as well as techniques such as randomization, bootstrapping, and simulation. (For more information about the CourseKata book and the principles guiding its design see Stigler et al., 2020; Son et al., 2021; Fries et al., 2021.)

In the mandatory once per week lab sessions, students were randomly paired with lab partners. Within each pair, roles were assigned randomly: one student as the performer, engaging directly in hands-on activities, and the other as the recorder, observing and recording the performer's hand movements using a smartphone. We hypothesized that there would be an interaction between the type of embodied intervention and the participants' prior knowledge. Specifically, we expected that performers with low prior knowledge would outperform recorders who also had low prior knowledge, whereas this advantage would not hold for those with high prior knowledge.

Methods

Participants

Participants were students enrolled in an introductory psychological statistics course at a large public research university. The class was structured as two weekly lectures and a lab session. The experimental interventions took place once a week for nine weeks during the 50-minute in-person lab sessions. Students received class participation points for attending lab sessions but were allowed to miss one lab session without losing any participation points.

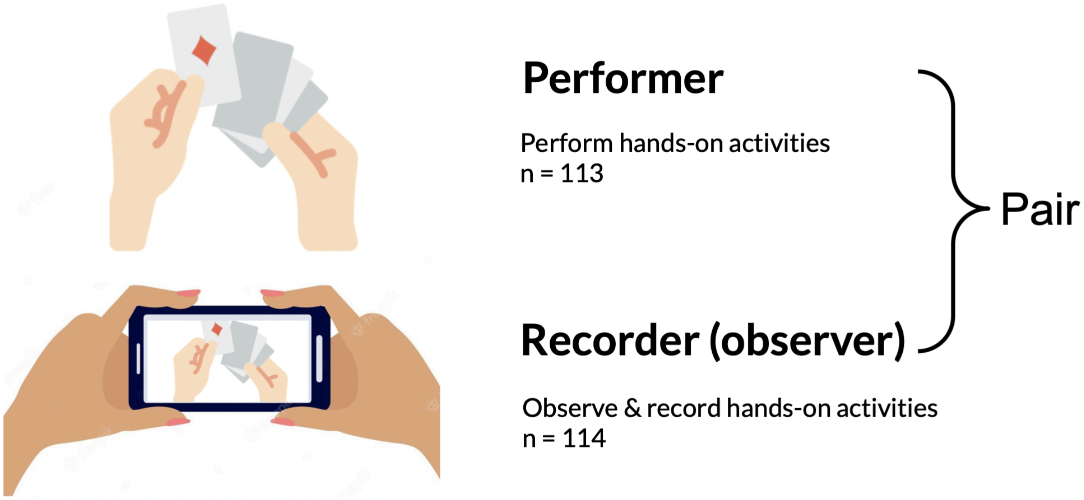
Although there were 236 students enrolled in the course, any students who did not agree to the data-sharing agreement of the course or who dropped the class were removed from the analysis. The final sample size was 227. Out of these students, 171 students self-identified as female (75%), 49 as male (22%), and 7 as non-binary (3%). The self-reported race and ethnicity of these students were as follows: 100 Asian or Asian American (44%), 8 Black or African American (4%), 47 Hispanic, Latino or Spanish origin (21%), 13 Middle Eastern or North African (6%), 46 White (20%), and 13 mixed/multi races (6%).

Design and Procedure

Participants were randomly assigned a lab partner at the beginning of the course. Within each student dyad, students were randomly assigned to either the perform condition (n = 113) or the observe/record condition (n = 114). Each student's role and partner stayed the same for the entire course (Figure 11).

Figure 11

Condition and Dyad



The participants assigned to the perform condition (i.e. the performers) were expected to perform embodied instructional actions designed to enhance learning. The participants assigned to the observe/record condition (i.e. the recorders) used their smartphone to video their partner's actions during the activity. If a student's partner could not make it to a lab, the student was temporarily paired with another student while maintaining their same role. We used a cover story to explain students' role assignment. Students were informed that we were interested in using student video data to design a pedagogical agent, and their recordings would provide valuable hand movement data.

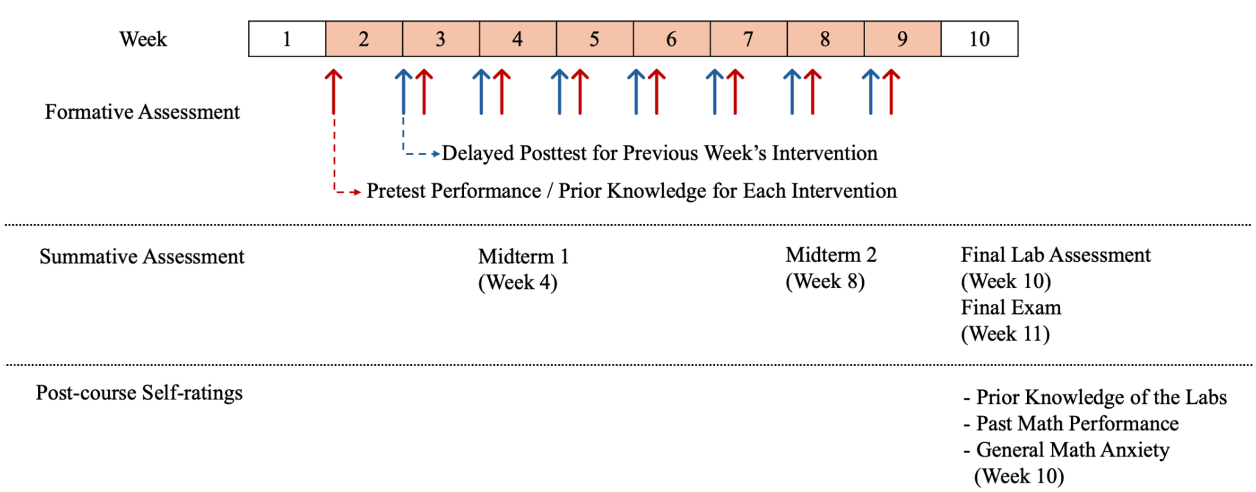
Starting from the second week, students began each lab by taking a set of "practice questions" on their laptops. They received a link to a Qualtrics survey and were given 6-10 minutes to complete the questions, with the time varying based on the number of questions. Half of these questions were designed as a delayed posttest assessment to measure what students had learned from the previous week. The other half were designed as a pretest to assess what students might already know about the concepts they were going to be taught in the current lab. (Because week 2 was the first lab activity, all questions from that week were pretest questions.)

After taking the assessments, students engaged in a lesson that incorporated embodied activities in the forms of object manipulation, gesture, and drawing. For example, students cut out a piece of paper with a small data table and used the pieces of paper to "shuffle" a variable and “resample” observations (i.e., sampling with replacement). In a later lab session, students used what they had learned about shuffle and resample in a hands-on activity in which they constructed sampling distributions. Based on their randomly assigned condition, students either performed these activities or observed and recorded their partner performing them.

During the course, there were two midterm exams. The first midterm was administered in week 4 and the second midterm was administered in week 8. At the end of the course, participants completed their final lab assessment and a lab exit survey in the 10th week and took a final exam in the 11th week. Figure 12 illustrates when the assessments and ratings were administered.

Figure 12

Timing of Each Assessment



Materials

A description of the lab schedule and activities is included in Appendix A.

Measures

Formative Assessments (Practice Questions)

Prior knowledge Assessments. Half of the practice questions administered to students at the beginning of each lab served as a pretest of prior knowledge related to the day's activities. A complete list of the prior knowledge assessment questions can be accessed on the study's OSF page (https://osf.io/ntsr2/?view_only=ce650267f407451a9ea26abcb428d8f7).

Delayed Post-Test Assessment. The other half of the practice questions answered at the start of each lab were designed as delayed post-test assessments of concepts worked on during the previous week's lab. (For a complete list of the questions see https://osf.io/ntsr2/?view_only=ce650267f407451a9ea26abcb428d8f7). .

Scoring of Formative Assessment Questions. Four trained coders graded participants' responses to the practice questions, with two coding each response as a means of assessing inter-rater reliability. All coding was conducted blind to condition.

The rubric the coders used to grade the responses was jointly determined in a meeting with the four coders and the lead researcher based on a preliminary analysis of a sample of responses. Each question was worth one point, with half points given to correct but incomplete answers or answers with minor misunderstandings. If the discrepancy rate between two coders on a question exceeded 20%, they would meet with the lead researchers to review the coding rubric and recode all responses to that question. If the discrepancy rate on a question was lower than 20%, the two coders would discuss and resolve their disagreements to arrive at a final score.

Summative Assessments

Final Lab Assessment. The final lab assessment was given to students in-person in week 10 at the last lab they were required to attend. It was composed of 28 questions (17 open-response, five multiple-choice, and six drawing questions). It was structured as a paper and pencil test. The same four coders scored students' responses using the same grading approach.

Midterm One & Midterm Two Performance. Midterm One consisted of 20 questions (15 multiple-choice and five open-response/coding questions). The exam assessed students' mastery of content covered in Chapters 1 to 5 in the book.

Midterm Two had 25 questions (18 multiple-choice and seven open-response/coding), which assessed students' mastery of content covered in Chapters 1 to 9 in the book.

The two midterms were graded by the two teaching assistants for the course, blind to the experimental condition students were assigned to. Each question was worth one point, with half points given to open-response questions that were incomplete or partially correct.

Final Exam. The final exam consisted of 44 questions designed to assess concepts covered in the entire book (Chapters 1 to 12). Out of the 44 questions, 29 were multiple choice and 15 were open response or coding questions. The same two teaching assistants graded the final exam following the same approach as their grading of the midterms.

Self-ratings at the End of the Course

Self-rated Overall Prior Knowledge of Lab Content. In addition to measuring students' prior knowledge of each separate lab we also measured students' overall prior knowledge of concepts covered in the lab interventions on an end-of-course survey at the end of week 10. We asked students to respond retrospectively to the question: "What percentage of concepts covered in the labs are concepts you already knew?" Students could choose 0%, 20%, 40%, 60%, 80% or 100%.

Past Math Performance. Previous math performance was self-rated by students at the end of the course on the same survey. Students rated how much they agreed with the statement, "My math/statistics marks were lower compared to other subjects", by choosing among five possible choices: strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree.

General Math Anxiety. Students' general math anxiety was measured at the same time as their past math performance. Students responded to the question, "On a scale of 1 to 6 (1 being not anxious at all and 6 being extremely anxious), how math anxious are you?"

Analyses

We tested our main hypothesis with two types of analyses. Analysis of CoVariance (ANCOVA) was used to examine whether students' self-ratings (i.e., their general prior knowledge, past math performance, and math anxiety) moderated the effect of condition on the summative assessments (the two midterms, final lab assessment, and final exam). We expected this interaction to be weak for the earlier summative assessments (e.g., midterm one) because the intervention had just begun, but stronger for assessments used later in the course.

Next, to investigate whether students' pretest performance moderated the effect of condition on the weekly formative delayed posttest assessments, we conducted a multilevel analysis with the repeated measures data. We expected a significant interaction between condition and pretest performance such that students with low pretest performance would benefit more from performing embodied pedagogies whereas students with high pretest performance would benefit more from observing embodied pedagogies.

Missing Data

Each student was assigned a partner ID to identify their lab partner. While the majority of students retained the same partner throughout the study, in instances where a student had more than one lab partner they were assigned the partner ID of the student with whom they were paired most frequently. When students missed a lab, or if their partner was switched during a lab, their data for that week (i.e., their corresponding pre and delayed posttest scores for that week) was treated as missing. Students who withdrew from the class midway, were not included in any data analyses.

Missing data were imputed using the Markov Chain Monte Carlo (MCMC) algorithm as implemented in the Blimp application (Enders, 2017).

Results

The deidentified data and data analysis syntax on which the present conclusions are based are available through the Open Science Framework (https://osf.io/ntsr2/?view_only=ce650267f407451a9ea26abcb428d8f7). The study design, hypotheses, and analytic plan were not pre-registered.

Analyses of Summative Assessments

The sample size for this part of the analysis is 217 (i.e., the number of students who completed the post survey). Because the final lab assessment was given in-person in their last structured lab session, there were a few more missing data for the final lab assessment (N = 195).

Interactions of Prior Knowledge and Condition

There were no significant main effects of condition on any of the summative assessments (midterm 1, midterm 2, the final exam, or the final lab assessment. This result was not changed by controlling for students' self-rated overall prior knowledge of the interventions (see Appendix B for the complete results).

Figure 13 shows the relationship between condition and self-rated prior knowledge predicting Midterm 1, Midterm 2, final exam, and final lab assessment. With the exception of midterm 1, all three other graphs showed a similar pattern of results: in the Observe condition, students' self-rated prior knowledge showed a positive correlation with their performance on the assessment, but such a correlation was not evident in the Perform condition.

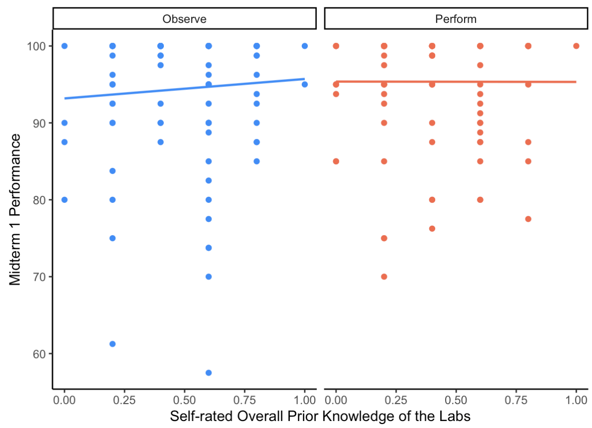
We tested whether there was a significant interaction between prior knowledge and condition for each of the four outcome variables using ANCOVA. Statistical analysis was carried out using R Studio (Version 2023.12.1.402, RStudio Team, 2023). Due to students dropping the class or not attending the session, four participants were removed from the analysis for midterm 1, midterm 2, and the final exam, and 26 participants were removed from the analysis for the lab final assessment.

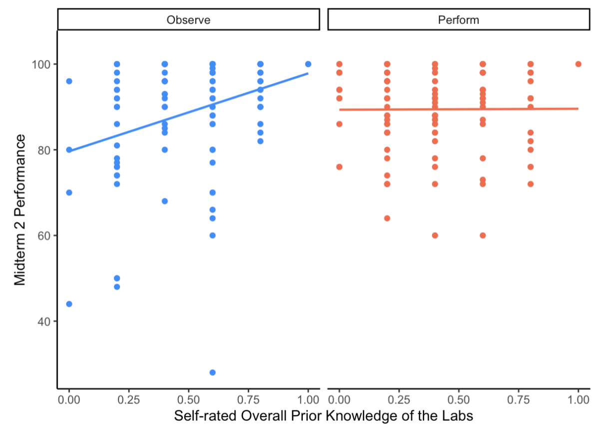
There was no significant interaction between self-rated prior knowledge of the labs and condition for midterm 1 ($F(1, 213) = 0.33$, $p = .564$). However, there was a significant interaction between students' self-rated prior knowledge and condition for midterm 2 ($F(1, 213) = 7.43$, $p = .007$), the final exam ($F(1, 213) = 8.41$, $p = .004$), and the final lab assessment ($F(1, 191) = 5.05$, $p = .026$). A complete table of results for the four models is included in Appendix C.

Figure 13

The Relationship Between Condition and Self-Rated Prior Knowledge for Each of the Four Summative Assessments

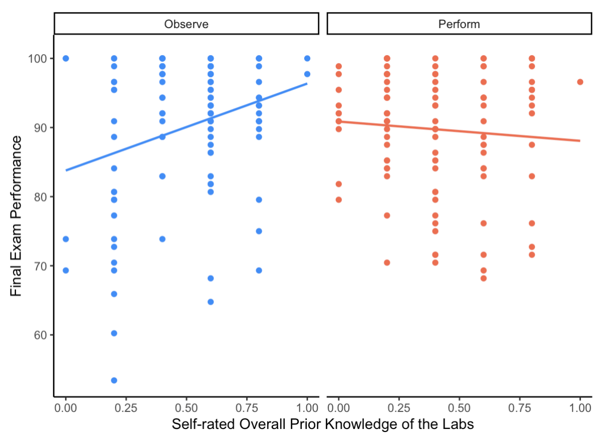
Midterm 1	Midterm 2

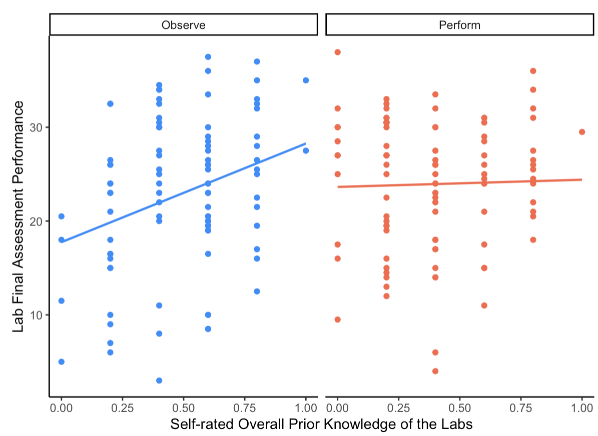




Final Exam

Final Lab Assessment

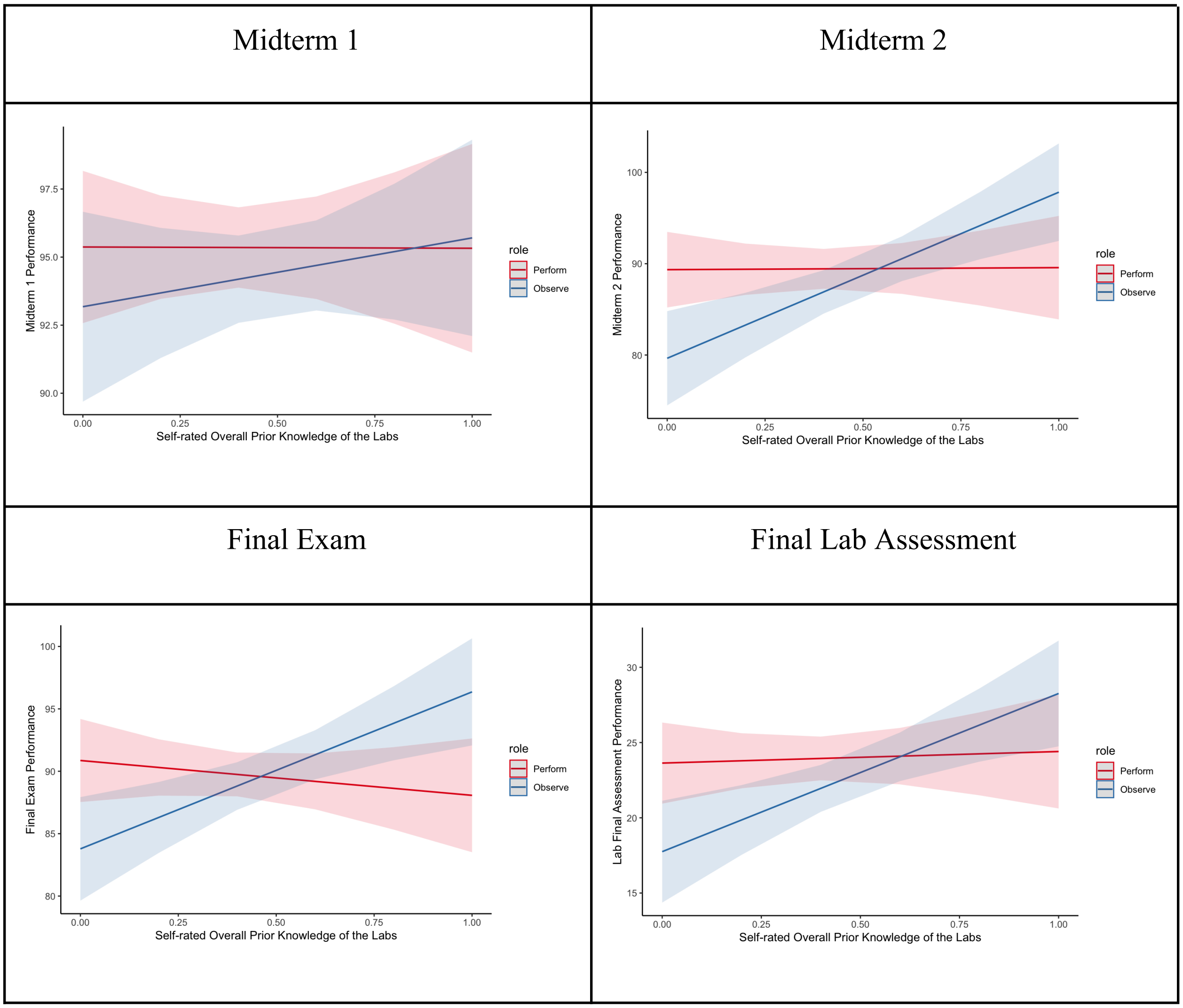




Model predictions and 95% confidence intervals are plotted in Figure 14, a suggested way to interpret significant interactions by Garofalo et al. (2022). From the graphs, we can see that for the three models where we detected significant interactions, the red and blue lines cross each other at approximately the 50% rating of prior knowledge. For students with lower self-rated prior knowledge, those in the Perform condition showed significantly superior learning in midterm 2, the final exam, and the final lab assessment than those in the Observe condition. (The shaded confidence interval does not contain the model prediction.) For students with higher self-rated prior knowledge, in contrast, those in the Observe condition achieved significantly better learning outcomes than students in the Perform condition (this difference is marginal in the final lab assessment).

Figure 14

Model by Self-Rated Prior Knowledge and Condition for each Summative Assessment

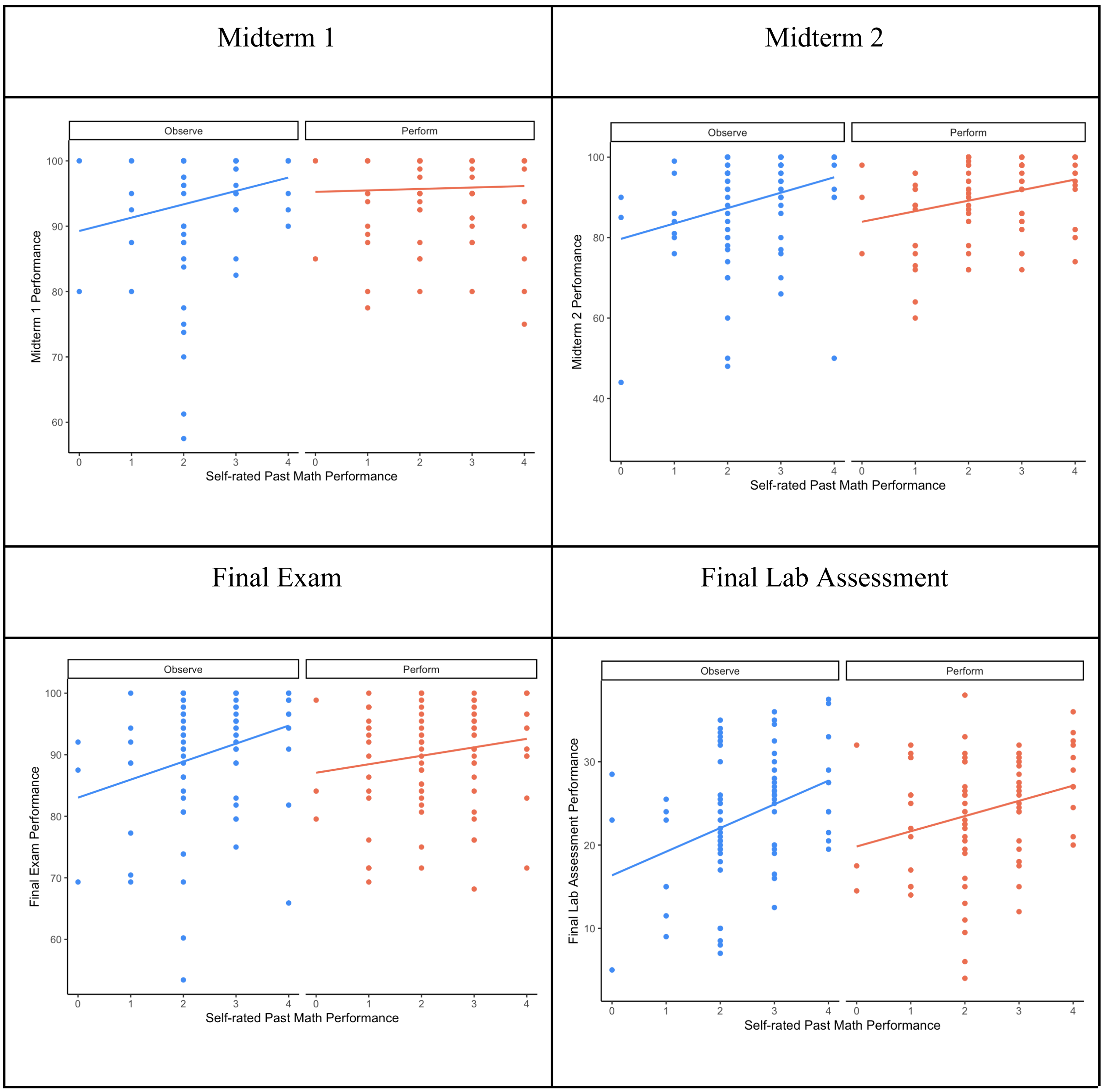


Interactions of Past Math Performance and Condition

Figure 15 shows the relationship between self-rated past math performance and the four summative assessments broken down by condition. Except for the graph for midterm 1, the other three graphs indicate a positive correlation between past math performance and outcome scores regardless of condition. In none of the summative assessments was the interaction between condition and self-rated past math performance significant.

Figure 15

The Relationship Between Condition and Self-rated Past Math Performance for Each of the Four Summative Assessments

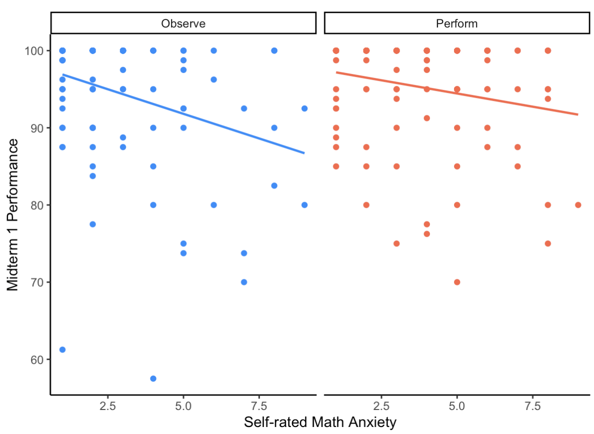
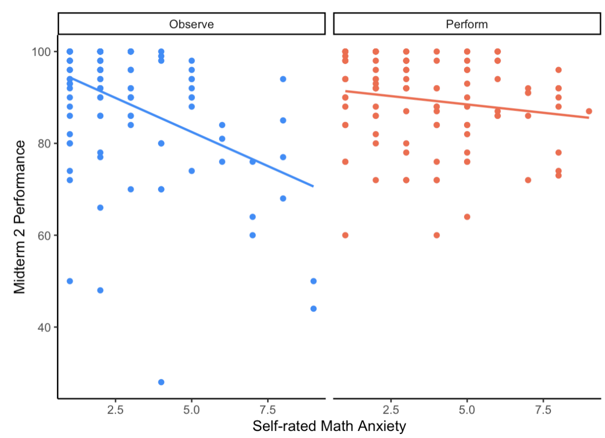


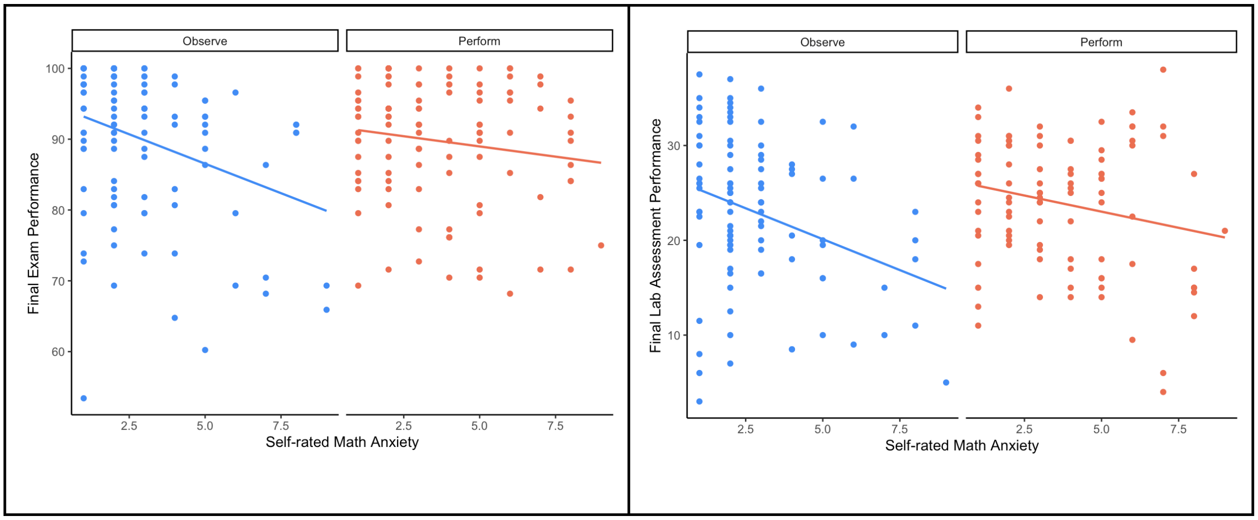
Interactions between Math Anxiety and Condition

Similarly, the relationship between students' self-rated math anxiety and each summative assessment separated by condition is shown in Figure 16. There was in general a negative relationship between math anxiety and performance regardless of the condition. Although the lines were steeper for the Observe group than the Perform group, we did not find any significant interaction between math anxiety and condition for any course assessment ($p > .05$).

Figure 16

The Relationship Between Condition and Self-rated Math Anxiety for Each of the Four Summative Assessments

<table>
<tr><td>Midterm 1</td><td>Midterm 2</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Final Exam</td><td>Final Lab Assessment</td></tr>
</table>



Analyses thus far provide support for the Performing First hypothesis. Students who rated themselves as lower in prior knowledge for the lab activities benefited more from performing relative to observing than did students with higher prior knowledge. But these analyses measure prior knowledge using students' self-ratings. We next use multilevel modeling to see if the same effects can be spotted using formative assessment data, collected prior to each lab, as a more targeted and objective measure of prior knowledge.

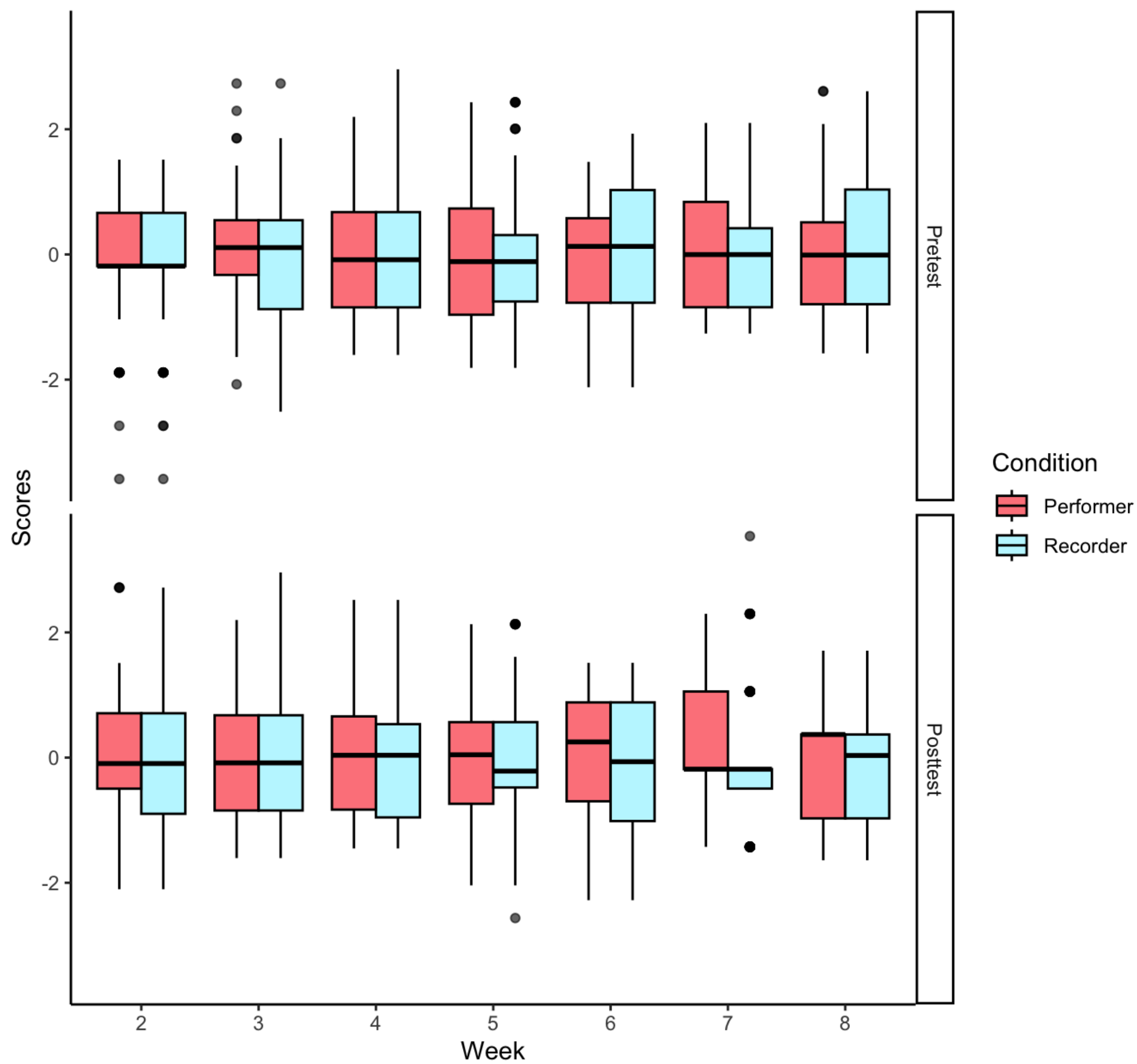
Analyses of Formative Assessments

Preprocessing and Plotting the Data

Because each week's pre- and post-test questions might differ in difficulty, we standardized within each test the scores prior to analysis so that each week's pre- and post-test scores are centered at 0. We plotted the standardized scores for each week to explore the differences between the two conditions in each week (Figure 17). From the graph, we can see that the Perform and Observe groups did not differ from each other on pretest questions. They also did not differ a lot from each other on the post-tests in the early weeks, but starting in week 5, the Perform group seemed to score a bit higher than the Observe group.

Figure 17

Standardized Pretest and Posttest Scores for each Week by Condition



Multilevel Model Specification

Given that the pre-test performance and post-test performance were not just collected once but repeatedly, there were three levels of analysis (i.e., repeated measures, student, and dyad). The sixteen repeated measures were nested within students and students were nested in dyads. Thus, we calculated the intraclass correlation coefficient (ICC) first for the post-test performance at each level followed by a similar analysis of pre-test performance. 68.8% of the total variability in post-test performance is due to repeated measures (level-1). Between-person mean differences (level-2) accounted for 23.8% of the variability in post-test performance. In other words, the expected correlation between two post-test scores of the same student is 23.8%, which is considered large in educational datasets. Lastly, between-dyad mean differences (level-3) accounted for 7% of the variability in post-test performance. While this percentage may seem relatively small, it was enough to warrant the inclusion of a third level for dyads in the model.

Similarly, for pre-test performance, repeated measures accounted for 77.6% of the total variability in pre-test scores. Between-person mean differences (level-2) accounted for 16.3% of the variability in pre-test performance. Lastly, between-dyad mean differences (level-3) accounted for 5.7% of the variability in pre-test performance.

We specified a three-level MLM. The first level was the repeated measures nested within students. The second level was the students, who are nested in dyads (i.e. the third level). Below, we show the equation for the overall model:

$$Posttest_{ijk} = \gamma_{000} + \gamma_{010}Condition_{jk} + \gamma_{020}Pretest_{jk}^{b.cgm} + \gamma_{030}Condition_{jk} * Pretest_{jk}^{b.cgm} + \gamma_{100}Pretest_{ijk}^{w} + \gamma_{110}Pretest_{ijk}^{w} * Condition_{jk} + \gamma_{200}Time_{ijk} + \gamma_{210}Condition_{jk} * Time_{ijk} + u_{00k} + r_{0jk} + e_{ijk}$$

In the equation above, we partitioned the variation of students' pretest performance into $Pretest_{jk}^{b.cgm}$, which stands for the mean of each student's pretest performance (i.e. the between-cluster variation) and $Pretest_{ijk}^{w}$, which stands for how each student's pretest performance varied from week to week (i.e. the within-cluster residual). The equations for each level are fully presented in Appendix D. Based on our hypothesis, we were particularly interested in γ_{03} and γ_{11}. Because the variation in pretest performance has both level-1 and level-2 variation, the interaction between pretest performance and condition (perform versus observe) was partitioned into the interaction between condition and each individual's average pretest performance (i.e. γ_{03}), which we will refer to as the level-2 interaction, and the interaction between condition and the variation within each individual's pretest performance (γ_{11}), which we will refer to as the cross-level interaction.

Interaction between Pretest and Condition

We used maximum likelihood estimation to fit the model using the Blimp application (Enders, 2017). There was no significant level-2 interaction between condition and pretest performance (median = 0.09, 95% credible interval = [-0.15, 0.34], but there was a significant cross-level interaction between condition and pretest (median = -0.12, 95% credible interval = [-0.22, -0.01]). A complete printout of the model results is included in Appendix E.

We next probed the significant cross-level interaction at the two levels of condition. For participants in the Observe condition, the correlation between pretest performance and delayed posttest performance was statistically significant (median = 0.11, 95% credible interval = [0.03, 0.18]). However, for participants in the Perform condition, the correlation between pretest performance and delayed posttest performance was not statistically significant (median = -0.01, 95% credible interval = [-0.08, 0.07]).

Discussion

In the current project, we designed and implemented an embodied lab curriculum in a college-level introductory statistics course taught using R programming. Central to our investigation were two questions: how to design and apply embodied pedagogies in a longitudinal academic setting and whether learners' prior knowledge moderates the effectiveness of these pedagogies. We proposed the Performing First Hypothesis, which posits that learners who have no or low prior knowledge of the concepts to be learned need to physically perform the activities to reap the most benefit out of the hands-on instruction, whereas learners with higher levels of prior knowledge can benefit similarly from simply observing a hands-on demonstration.

In two sets of analyses, we demonstrated the pivotal role of prior knowledge in moderating the effect of performing versus observing actions and gestures. First, we found that learners' self-ratings (at the end of the course) of their overall prior knowledge of the concepts covered in the labs significantly moderated the effect of condition on their midterm 2, final exam, and final lab assessment performance. We did not find self-rated prior knowledge to moderate the effect of condition on the first midterm, for which there are three potential explanations. First, there may be a ceiling effect in students' performance on midterm 1. Second, the overall prior knowledge rating may not accurately capture students' prior knowledge in the first weeks of the course. Third, and more related to the investigation of why embodied pedagogies are important in learning, is that the absence of this moderating effect on the first midterm can possibly be attributed to the students having experienced only two lab sessions by that point.

Students might need to practice performing actions and gestures more than two times in order to construct mental representations of concepts that are robust enough to be connected to other concepts. In weeks 2 and 3 (i.e., the two weeks before the midterm) students learned to enact or observe hands-on shuffling versus resampling of a dataset. In the later weeks of the course, students used what they had learned about shuffle and resample to engage in hands-on activities related to creating and manipulating sampling distributions. In this way, the mental representations they acquired in earlier weeks were further developed and connected. This is one of the strengths of conducting an intervention longitudinally because such an approach allows us to detect effects that would not have been detected in an intervention that lasted only two sessions.

We also found that other factors, such as students' previous mathematics performance and mathematics anxiety, did not significantly moderate the effect of condition. This suggests that what matters is not whether learners had more positive experiences with mathematics in the past, but whether they came into the course with already-developed mental representations related to the concepts that were the focus of the intervention. Past literature has also highlighted how actions can lead to conceptual insights and the development of more robust mental representations (Brooks et al., 2018; Broaders et al., 2007; Cartmill et al., 2012; Goldin-Meadow et al., 2009). Our findings contribute to this literature by positing a unique role of learners' prior knowledge and putting forward the Performing First Hypothesis that serves as a framework for such investigations.

Under the framework of the Performing First Hypothesis, it is hypothesized that prior knowledge of the concepts to be learned—and not other factors—would moderate the impact of varying levels of embodiment (i.e., performing versus observing) on learning outcomes. Our results, which revealed a significant interaction consistent with this hypothesis, validate the proposed mechanism and support the theory that embodied pedagogies differentially influence learners. Future research should deliberately seek evidence to further elucidate this mechanism, for instance, by exploring the type and quality of mental representations formed under various embodied and non-embodied pedagogical approaches among diverse learners. Preliminary data from our research, currently under review, indicate that more embodied pedagogies foster enhanced visuospatial representations of the concepts taught compared to both less embodied and non-embodied approaches. These mental representations appear to mediate the influence of differing pedagogical strategies on students' learning outcomes (Authors et al., under review).

Our measure of students' overall prior knowledge of the lab interventions is retrospective and self-reported. It is possible that students' rating of prior knowledge is impacted by their experience in the intervention instead of the other way around. This is why we supplemented our analysis with the MLM, which used students' actual performance on each week's pretest to measure their prior knowledge. Consistent with our Performing First hypothesis and previous analysis, the three-level MLM revealed a significant cross-level interaction between the type of embodied intervention and learners' prior knowledge. Specifically, when learners observed hands-on activities, there was a significant correlation between their prior knowledge and their performance on a delayed posttest. Conversely, when learners engaged in physical activities themselves, this correlation between prior knowledge and posttest performance was not evident. This suggests that while prior knowledge typically predicts post-intervention performance, active participation in physical tasks disrupts this predictive relationship, possibly due to the different cognitive processes involved in physical versus observational learning. The de-emphasis of prior knowledge when learners engage in active performance delivers a powerful message to educators, especially from an equity perspective.

The findings from the current study not only provide support for the Performing First Hypothesis but also suggest new directions for embodied learning research. Past research on embodied learning has demonstrated the efficacy of embodied interventions (i.e., interventions that leverage bodily experience in some form) when compared with abstract instructions that do not involve the body. The current study suggests more nuanced questions of what and when - what type of embodied pedagogies are the most effective, and when are they the most effective?

In the current study, we focused on the distinction between performing and observing hands-on activities, as both forms of embodied pedagogy are prevalent in both research and practice. We hypothesized that there is something inherently unique about "performing" activities, a concept that is supported by classical developmental psychology. Piaget, for example, highlighted the sensorimotor stage as foundational for later higher-order thinking and abstract concept processing (Piaget, 1983; Piaget & Inhelder, 1969). Nonetheless, we also recognize the potential value in exploring other types of embodied pedagogies, including whole-body movements, virtual reality (VR) manipulations, and directed actions, as these too may offer significant educational benefits.

We are not the only researchers who have started to pay attention to the role prior knowledge plays in learners' response to embodied pedagogies (Cook et al., 2024, Congdon et al., 2018; Congdon & Goldin-Meadow, 2021). For example, in addition to the study by Congdon et al., 2018 reviewed in the introduction, a recent study by Cook and colleagues (2024) randomly assigned second- and third-grade students to watch an instructional video about mathematical equivalence that either included gestures or not. The researchers found a complex interaction between learners' prior knowledge (as demonstrated by their pretest strategy), the condition (gesture vs. no gesture), the question type (whether the question was a conceptual question or procedural question) and the way they asked the question (whether there was interference from prior knowledge). Although it is difficult to interpret the exact meaning of the four-way interaction, the findings suggest that the effect of gesture depends on the learners' prior knowledge and the nature of the questions.

These are just the beginning of such an investigation to explore the intricate relationship between prior knowledge and different types of embodied pedagogies. Our Performing First theory provides a framework for forthcoming research, which will allow researchers to generate testable hypotheses and experiments that are crucial to moving the field forward. We urge future studies to investigate the relationship between prior knowledge and other forms of embodied pedagogies to fully elucidate the question of “what” types of pedagogies and “when” in learners’ knowledge development would embodied pedagogies be beneficial.

Lastly, these findings have important implications for instructional design in STEM education. Although teacher demonstrations are much easier to implement in real classrooms, for novices it seems more beneficial to incorporate active participation in physical performance into the curriculum to facilitate a deeper understanding of the concepts. After students have gained meaningful hands-on experience with the concepts, teachers can then switch to teacher demonstrations or perhaps even more abstract instruction. This research highlights the importance of tailoring educational strategies to the learner's prior knowledge. Future research should continue exploring the nuanced ways in which different types of embodied pedagogies impact learners at different time points of their knowledge development. This could lead to more refined strategies and curriculums that cater to the diverse needs of learners in higher education, particularly in STEM domains.

Appendix A: lab schedule

Week	**Content**
Week 1	General introduction
Week 2	Cutting out and shuffling a dataset to understand shuffle ()
Week 3	Cutting out and sampling with replacement to understand resample()
Week 4	Comparing shuffle() and resample()
Week 5	Drawing and gesturing to understand the empty model and the two-group model
Week 6	Drawing and gesturing to understand the three-group model
Week 7	Drawing and gesturing to understand the regression model, and compare it with group models
Week 8	Using shuffle () and resample () to simulate sampling distributions
Week 9	Hands-on construction of the confidence interval
Week 10	Lab general assessment and debrief

Appendix B

Model: midterm1 ~ role + concept_known

Predictors	*Estimates*	*df*	*F*	*CI*	*p*
(Intercept)	93.86			91.25 – 96.47	**<0.001**
role [Perform]	1.04	1	0.909	-1.11 – 3.18	0.342
concept known	1.13	1	0.261	-3.24 – 5.50	0.610
Observations	217				
R^2 / R^2 adjusted	0.005 / -0.005				

Model: midterm2 ~ role + concept_known

Predictors	*Estimates*	*df*	*F*	*CI*	*p*
(Intercept)	84.41			80.48 – 88.34	**<0.001**
role [Perform]	1.63	1	0.992	-1.59 – 4.85	0.320
concept known	8.45	1	6.426	1.88 – 15.02	**0.012**
Observations	217				
R^2 / R^2 adjusted	0.030 / 0.021				

Model: final exam ~ role + concept_known

Predictors	*Estimates*	*df*	*F*	*CI*	*p*
(Intercept)	87.86			84.70 – 91.03	**<0.001**
role [Perform]	0.17	1	0.017	-2.42 – 2.77	0.896

concept known	4.25	1	2.500	-1.05 – 9.54	0.115
Observations	217				
R^2 / R^2 adjusted	0.012 / 0.002				

Model: final lab assessment ~ role + concept_known

Predictors	*Estimates*	*df*	*F*	*CI*	*p*
(Intercept)	20.32			17.75 – 22.89	**<0.001**
role [Perform]	1.55	1	2.061	-0.58 – 3.68	0.153
concept known	5.25	1	5.780	0.94 – 9.56	**0.017**
Observations	195				
R^2 / R^2 adjusted	0.034 / 0.024				

Appendix C

Model: midterm1 ~ role * concept_known

Predictors	*Estimates*	*df*	*F*	*CI*	*p*
(Intercept)	93.18			89.67 – 96.68	**<0.001**
role [Perform]	2.19	1	0.926	-2.30 – 6.69	0.337
concept known	2.53	1	0.594	-3.94 – 8.99	0.442
role [Perform] * concept known	-2.57	1	0.334	-11.36 – 6.21	0.564
Observations	217				
R^2 / R^2 adjusted	0.006 / -0.008				

Model: midterm2 ~ role * concept_known

Predictors	*Estimates*	*df*	*F*	*CI*	*p*
(Intercept)	79.64			74.46 – 84.83	**<0.001**
role [Perform]	9.70	1	8.281	3.06 – 16.35	**0.004**
concept known	18.19	1	14.047	8.62 – 27.75	**<0.001**
role [Perform] * concept known	-17.97	1	7.430	-30.96 – -4.97	**0.007**
Observations	217				
R^2 / R^2 adjusted	0.063 / 0.050				

Model: final exam ~ role * concept_known

Predictors	*Estimates*	*df*	*F*	*CI*	*p*
(Intercept)	83.78			79.61 – 87.95	**<0.001**
concept known	12.58	1	6.821	4.89 – 20.27	**0.001**
role [Perform]	7.08	1	10.391	1.74 – 12.43	**0.010**
concept known * role [Perform]	-15.37	1	8.410	-25.82 – -4.92	**0.004**
Observations	217				
R^2 / R^2 adjusted	0.049 / 0.036				

Model: final lab assessment ~ role * concept_known

Predictors	*Estimates*	*df*	*F*	*CI*	*p*
(Intercept)	17.75			14.35 – 21.15	**<0.001**
concept known	10.51	1	7.120	4.23 – 16.80	**0.001**
role [Perform]	5.89	1	10.882	1.54 – 10.24	**0.008**
concept known * role [Perform]	-9.74	1	5.048	-18.30 – -1.19	**0.026**
Observations	195				
R^2 / R^2 adjusted	0.059 / 0.044				

Appendix D: equation for each level

Level 1:

$$Posttest_{ijk} = \pi_{0jk} + \pi_{1jk}(Pretest_{ijk}^{w}) + \pi_{2jk}Time_{ijk} + e_{ijk}$$

Level 2:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}Condition_{jk} + \beta_{02k}Pretest_{jk}^{b.cgm} + \beta_{03k}Condition_{jk} * Pretest_{jk}^{b.cgm} + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k} + \beta_{11k}Condition_{jk}$$

$$\pi_{2jk} = \beta_{20k} + \beta_{21k}Condition_{jk}$$

Level 3:

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

$$\beta_{01k} = \gamma_{010}$$

$$\beta_{02k} = \gamma_{020}$$

$$\beta_{03k} = \gamma_{030}$$

$$\beta_{10k} = \gamma_{100}$$

$$\beta_{11k} = \gamma_{110}$$

$$\beta_{20k} = \gamma_{200}$$

$$\beta_{21k} = \gamma_{210}$$

Appendix E: Results of the Three-level MLM

Coefficients	Median	StdDev	2.5% CI	97.5% CI
Intercept	0.039	0.096	-0.150	0.228
Condition.Exp	-0.100	0.129	-0.350	0.154
Pretest	0.107	0.037	0.034	0.181
Time	-0.022	0.015	-0.052	0.008
Pretest.mean[ID]	1.062	0.139	0.794	1.342
Pretest.mean[Dyad]	-0.003	0.312	-0.593	0.631
Condition.Exp*Pretest	-0.114	0.053	-0.218	-0.012
Condition.Exp*Time	0.044	0.022	0.001	0.086
Condition.Exp*Pretest. mean[ID]	0.178	0.187	-0.180	0.556
Condition.Exp*Pretest. mean[Dyad]	-0.270	0.417	-1.103	0.549

* CI stands for the credible interval.

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 Media SA.

Authors et al. (Under Review). Watching Hands Move Enhances Learning from Concrete and Dynamic Visualizations.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, *22*(4), 577-660.

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

Borghi, A. M., & Pecher, D. (2011). Introduction to the special topic embodied and grounded cognition. *Frontiers in Psychology*, *2*, 187.

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.

Brod, G. (2021). Toward an understanding of when prior knowledge helps or hinders learning. *npj Science of Learning*, *6*(1), 24.

Brooks, N. B., Barner, D., Frank, M., & Goldin‐Meadow, S. (2018). The role of gesture in supporting mental representations: The case of mental abacus arithmetic. *Cognitive Science*, *42*(2), 554-575.

Calvo-Merino, B. (2013). Neural mechanisms for action observation. *People watching: Social, perceptual, and neurophysiological studies of body perception*, 283-303.

Cartmill, E. A., Beilock, S., & Goldin-Meadow, S. (2012). A word in the hand: action, gesture and mental representation in humans and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 129-143.

Chakraborty, J., & Esposito, A. G. (2024). Adult Learners Self‐Derive New Knowledge through Integration of Novel Information and Prior Knowledge and Are more Successful with Reactivation. *Mind, Brain, and Education*.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. OUP USA.

Congdon, E. L., & Goldin-Meadow, S. (2021). Mechanisms of embodied learning through gestures and actions: Lessons from development. *Handbook of embodied psychology: Thinking, feeling, and acting*, 527-546.

Congdon, E. L., Kwon, M. K., & Levine, S. C. (2018). Learning to measure through action and gesture: Children's prior knowledge matters. *Cognition*, *180*, 182-190.

Cook, S. W., Wernette, E. M., Valentine, M., Aldugom, M., Pruner, T., & Fenn, K. M. (2024). How Prior Knowledge, Gesture Instruction, and Interference After Instruction Interact to Influence Learning of Mathematical Equivalence. *Cognitive Science*, *48*(2), e13412.

Enders, C. K. (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. *Behaviour research and therapy*, *98*, 4-18.

Fries, L., Son, J. Y., Givvin, K. B., & Stigler, J. W. (2021). Practicing connections: A framework to guide instructional design for developing understanding in complex domains. *Educational Psychology Review*, *33*(2), 739-762.

Fu, Y., & Franz, E. A. (2014). Viewer perspective in the mirroring of actions. *Experimental brain research*, *232*, 3665-3674.

Garofalo, S., Giovagnoli, S., Orsoni, M., Starita, F., & Benassi, M. (2022). Interaction effect: Are you doing the right thing?. *PLoS One*, *17*(7), e0271668.

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage*, *35*(4), 1674-1684.

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in cognitive sciences*, *9*(5), 234-241.

Gerson, S. A., & Woodward, A. L. (2014). Learning from their own actions: The unique effect of producing actions on infants' action understanding. *Child development*, *85*(1), 264-277.

Grotelueschen, A. D. (1979). Introductory material structure and prior knowledge effects on adult learning. *Adult Education*, *29*(2), 75-82.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological science*, *20*(3), 267-272.

Goldin‐Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental science*, *15*(6), 876-884.

Golonka, S., & Wilson, A. D. (2012). Gibson's ecological approach. *Avant: Trends in Interdisciplinary Studies*, *3*(2), 40-53.

Guarino, K. F., & Wakefield, E. M. (2020). Teaching analogical reasoning with co-speech gesture shows children where to look, but only boosts learning for some. *Frontiers in Psychology*, *11*, 575628.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of educational psychology*, *106*(1), 86.

Johnson-Glenberg, M. C., & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive research: principles and implications*, *2*(1), 1-28.

Lau, W. W. F. & Yuen, A. H. K. (2008). Exploring the effects of gender and learning styles on computer programming performance: Implications for programming pedagogy. British Journal of Educational Technology, doi: 10.1111/j.1467-8535.2008.00847.x

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

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological science*, *25*(4), 903-910.

Novack, M., & Goldin-Meadow, S. (2015). Learning from gesture: How our hands change our minds. *Educational psychology review*, *27*, 405-412.

Pezzulo, G., & Calvi, G. (2011). Computational explorations of perceptual symbol systems theory. *New Ideas in Psychology*, *29*(3), 275-297.

Piaget, J. (1983). Piaget's theory. *Handbook of child psychology*, *1*(4).

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.*, *27*, 169-192.

Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., & van der Wel, R. (2012). Cognition, action, and object manipulation. *Psychological bulletin*, *138*(5), 924.

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational psychologist*, *57*(1), 31-54.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1-B11.

Sommerville, J. A., & Woodward, A. L. (2010). The link between action production and action processing in infancy. *Naturalizing intention in action*, 67-89.

Son, J. Y., & Stigler, J. W. (2024). Statistics and data science: A modeling approach . https://coursekata.org/

Stigler, J. W., Son, J. Y., Givvin, K. B., Blake, A. B., Fries, L., Shaw, S. T., & Tucker, M. C. (2020). The Better Book approach for education research and development. *Teachers College Record*, *122*(9), 1-32.

Sullivan, J. V. (2018). Learning and embodied cognition: A review and proposal. *Psychology Learning & Teaching*, *17*(2), 128-143.

Thompson, R. A., & Zamboanga, B. L. (2003). Prior knowledge and its relevance to student achievement in introduction to psychology. *Teaching of Psychology*, *30*(2), 96-101.

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research*, 64, 37–54. https://doi.org/10.3102/00346543064001037

Wakefield, E. M., & James, K. H. (2015). Effects of learning with gesture on children's understanding of a new language concept. *Developmental Psychology*, *51*(8), 1105.

West, M., & Ross, S. (2002). Retaining females in computer science: A new look at a persistent problem. JCSC, 17(5), 1-7.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, *9*, 625-636.

Zacharia, Z. C., Loizou, E., & Papaevripidou, M. (2012). Is physicality an important aspect of learning through science experimentation among kindergarten students?. *Early Childhood Research Quarterly*, *27*(3), 447-457.

Zhang, I., Givvin, K. B., Sipple, J. M., Son, J. Y., & Stigler, J. W. (2021). Instructed hand movements affect students' learning of an abstract concept from video. *Cognitive science*, *45*(2), e12940.

Zhang, I. Y., Tucker, M. C., & Stigler, J. W. (2022). Watching a hands-on activity improves students' understanding of randomness. *Computers & Education*, *186*, 104545.

Zhang, I. Y., Guo, X. H., Son, J. Y., Blank, I. A., & Stigler, J. W. (2024). Watching videos of a drawing hand improves students' understanding of the normal probability distribution. *Memory & Cognition, 1-20.*