Computers & Education 186 (2022) 104545

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Watching a hands-on activity improves students' understanding of randomness☆

Icy (Yunyi) Zhang*, Mary C. Tucker, James W. Stigler

Department of Psychology, University of California, Los Angeles, USA

ARTICLE INFO

Keywords:
Hands-on demonstration
Computer simulation
Statistics education
Multimedia learning
Online instruction
Instructional sequence
Embodied cognition

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.

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.

☆ This project has been made possible in part by a grant from the Chan Zuckerberg Foundation and a grant from the California Governor's Office of Planning and Research (OPR18115).

* Corresponding author. UCLA Psychology, 90095-1563, USA.
E-mail address: yunyi9847@g.ucla.edu (I.(Y. Zhang).

https://doi.org/10.1016/j.compedu.2022.104545
Received 13 January 2022; Received in revised form 3 May 2022; Accepted 6 May 2022
Available online 13 May 2022

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.

1. 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.

2. Study 1

2.1. Method

2.1.1. Participants

Thirty-three undergraduate students from Uiversity of California, Los Angeles participated in the study. All students had completed the same introductory statistics course in the psychology department, taught by two different instructors, during the previous academic quarter. Both instructors used the same online textbook: *CourseKata Statistics and Data Science: A Modeling Approach* (Son & Stigler, 2017-2022). 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.

2.1.2. 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).

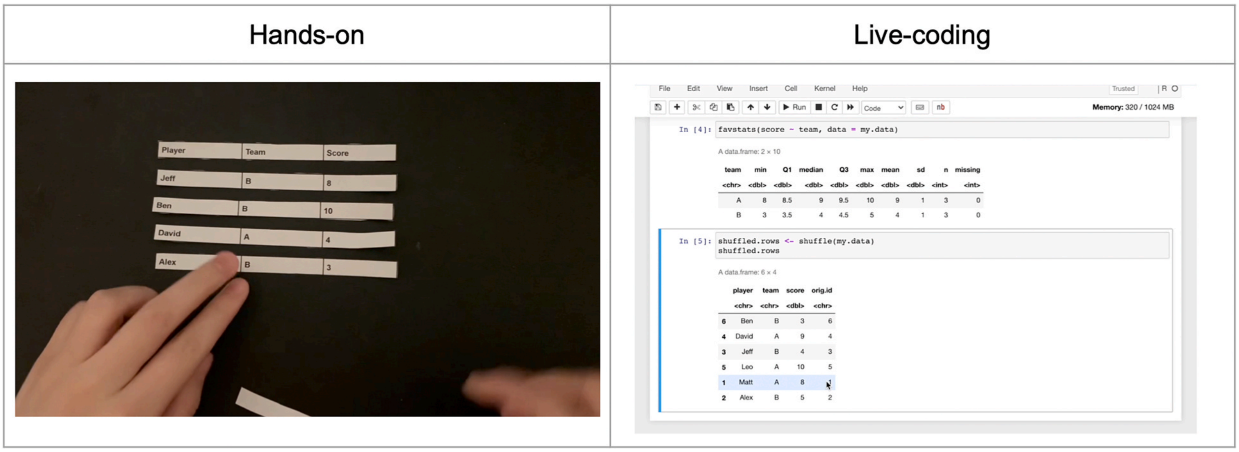
2.1.3. Materials

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. (Fig. 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.

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.



```
In [4]: favstats(score ~ team, data = my.data)

A data.frame: 2 × 10
 team   min    Q1  median    Q3   max  mean    sd     n missing
<chr> <dbl> <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <int>   <int>
    A     8   8.5       9   9.5    10     9     1     3       0
    B     3   3.5       4   4.5     5     4     1     3       0

In [5]: shuffled.rows <- shuffle(my.data)
        shuffled.rows

A data.frame: 6 × 4
  player  team score orig.id
   <chr> <chr> <dbl>   <chr>
6    Ben     B     3       6
4  David     A     9       4
3   Jeff     B     4       3
5    Leo     A    10       5
1   Matt     A     8       1
2   Alex     B     5       2
```

Fig. 1. Screen grabs from the hands-on video and the live-coding video.

2.1.4. *Measures*

2.1.4.1. 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 0–10, 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.

2.1.4.2. 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 0–10, 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."

3. Results

An analysis of pretest scores found no significant difference across conditions in students' prior understanding of the shuffle function ($t\,(31) = 0.17$, $\eta^2 = 0.00$, 90% CI = [0.00, 0.06], $p = .864$).

Fig. 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 ($t\,(31) = 2.27$, $\eta^2 = 0.14$, 90% CI = [0.01, 0.34], $p = .031$). Similar benefits of the hands-on group were observed when pretest was included as a covariate in the multiple linear regression model ($t\,(30) = 2.23$, $\eta_p^2 = 0.15$, 90% CI = [0.01, 0.34], $p = .033$). When included as a covariate (i.e., controlling for condition), students' pretest performance did not predict posttest scores ($t\,(30) = 0.86$, $\eta_p^2 = 0.02$, 90% CI = [0.00, 0.21], $p = .396$).

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 ($t\,(31) = 2.35$, $\eta^2 = 0.15$, 90% CI = [0.00, 0.38], $p = .025$); 2) describe what a specific line of code that shuffled the outcome variable in the dataset was doing ($t\,(31) = 2.06$, $\eta^2 = 0.12$, 90% CI = [0.01, 0.35], $p = .048$). 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 ($t\,(31) = 1.96$, $\eta^2 = 0.11$, 90% CI = [0.00, 0.34], $p = .059$).

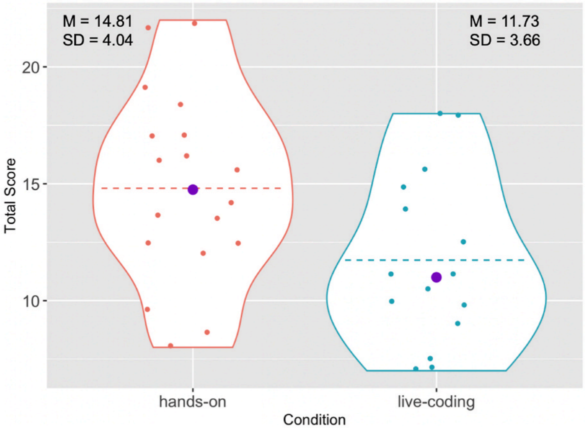


Fig. 2. Violin Plot Showing Posttest Scores by Condition. *Note.* Dashed lines show the mean of each group. Purple dots show the median. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

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, $\eta^2 = 0.05$, 90% CI = [0.00, 0.23], $p = .204$). 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) = 0.40, $\eta^2 = 0.00$, 90% CI = [0.00, 0.13], $p = .691$).

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, $\eta^2 = 0.12$, 90% CI = [0.00, 0.35], $p = .049$). However, students' change in self-rated understanding from pre to post intervention did not significantly correlate with performance (t (31) = 1.29, $\eta^2 = 0.05$, 90% CI = [0.00, 0.26], $p = .207$).

4. 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.

5. Study 2

5.1. Method

5.1.1. 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 0.14, obtaining a power of .7 or 0.8 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 0.7 and 0.8. As before, the study design and procedures were approved by the institutional review board for protection of human subjects.

5.1.2. 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.

6. 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 ($t\,(45) = -0.07$, $\eta^2 = 0.00$, 90% CI = [0.00, 0.00], $p = .945$).

Fig. 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 ($t\,(45) = 2.28$, $\eta^2 = 0.10$, 90% CI = [0.01, 0.26], $p = .028$). As in Study 1, this difference remained statistically significant when controlling for students' performance on the two-question pretest using a multiple linear regression ($t\,(44) = 2.35$, $\eta_p^2 = 0.11$, 90% CI = [0.01, 0.27], $p = .023$).

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 ($t\,(45) = 3.80$, $\eta^2 = 0.24$, 90% CI = [0.06, 0.44], $p < .001$). 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 ($t\,(45) = 2.96$, $\eta^2 = 0.16$, 90% CI = [0.02, 0.36], $p = .005$); 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 ($t\,(45) = 2.61$, $\eta^2 = 0.13$, 90% CI = [0.01, 0.32], $p = .012$).

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 ($t\,(31) = 1.39$, $\eta^2 = 0.04$, 90% CI = [0.00, 0.18], $p = .173$), nor did their ratings of how much they would like to see more activities like this in the future ($t\,(40) = 1.26$, $\eta^2 = 0.04$, 90% CI = [0.00, 0.21], $p = .216$). Also as in Study 1, linear regressions showed that students' post-intervention ratings of understanding significantly predicted performance on the posttest ($t\,(41) = 2.54$, $\eta^2 = 0.14$, 90% CI = [0.00, 0.34], $p = .015$), whereas participants' change in self-rated understanding from pre to post intervention did not significantly predict posttest performance ($t\,(41) = 0.19$, $\eta^2 = 0.00$, 90% CI = [0.00, 0.08], $p = .851$).

7. 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

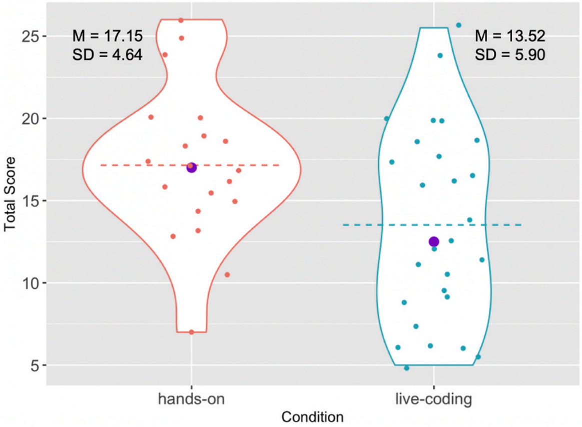


Fig. 3. Participants' Performance on Posttest by Condition. *Note.* Dashed lines show the mean of each group. Purple dots show the median. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

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.

7.1. 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 sequences.

Credit author statement

Icy (Yunyi) Zhang: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. Mary C. Tucker: Conceptualization, Methodology, Writing - Review & Editing. James W. Stigler: Conceptualization, Methodology, Writing - Review & Editing, Supervision, Funding acquisition.

Declaration of competing interest

We have no conflict of interest to disclose.

Appendix A

Pre-Test Attitudinal Measures.

1. In Psych 100 A, 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 100 A 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.

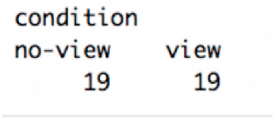
The laptop_data dataset contains data from an experiment on the effect of laptops on student learning. Undergraduate students were randomly assigned to one of two conditions: view or no-view. In the view condition, students attended a 40 min lecture and were allowed to keep their laptops open. In the 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)

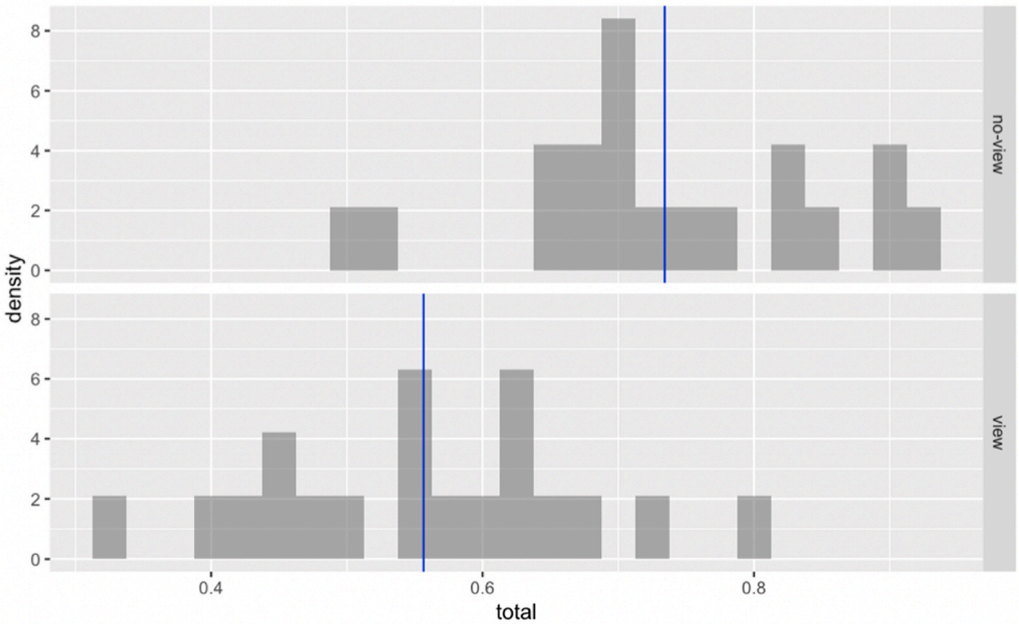


```
condition
no-view    view
     19      19
```

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:

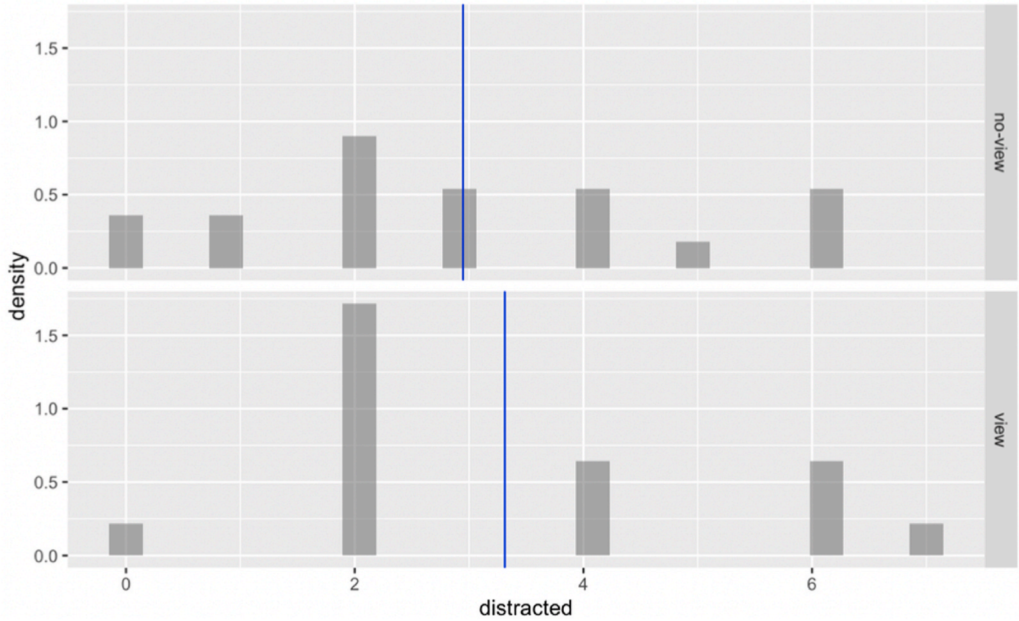
```
gf_dhistogram (∼ distracted, data = laptop_data) %>%
gf_facet_grid (shuffle (condition) ∼ .)
```

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 ∼ .)
```

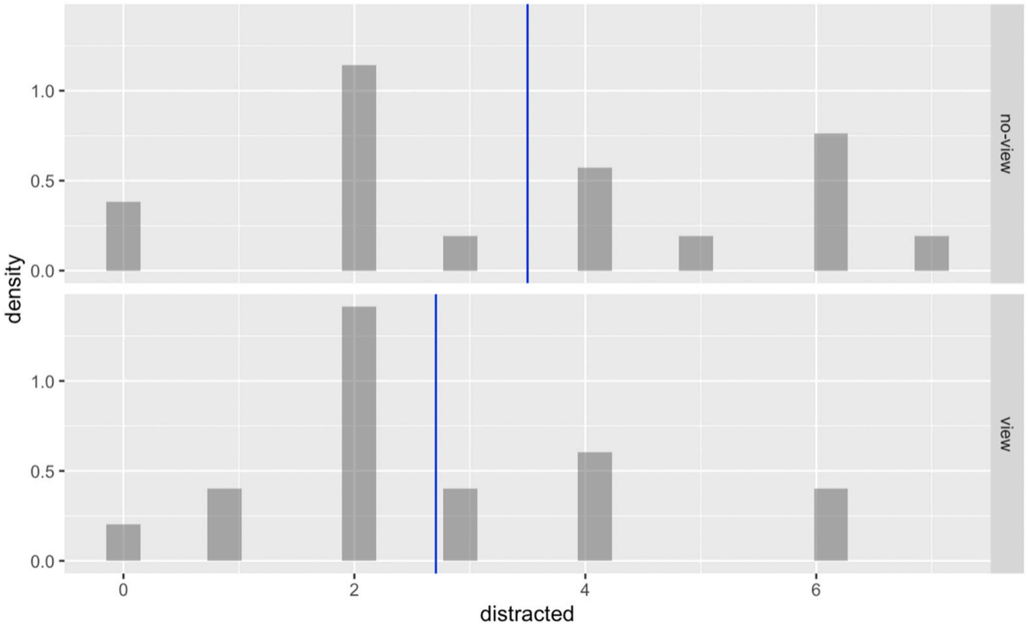


12. Sometimes groups differ just because of randomness. Do you think the group difference in the histogram above could be due to randomness?
    a. Yes, it must be due to randomness
    b. No, it cannot be due to randomness
    c. Maybe, we need to further investigate
13. If you ran the code in the previous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

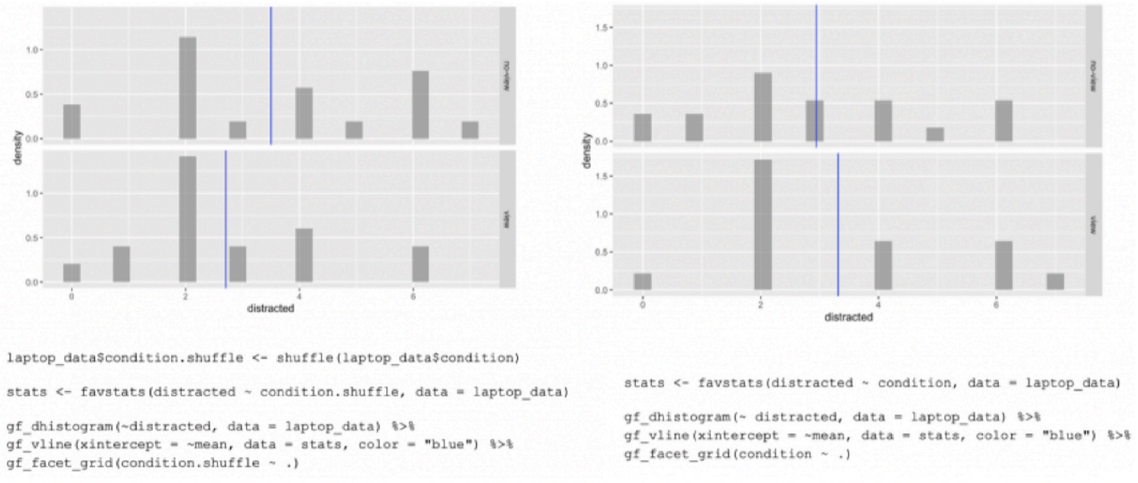
14. What caused the difference in the means represented in the graphs below?

```
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:



```
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 ~ .)
```

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 1–10, how math anxious are you?
2. In Psych 100 A, you learned how to do some R programming. On a scale of 1–6 (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 100 A class?
   a. Yes
   b. No
   c. Not sure/can't remember
4. On a scale of 1–10, 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 min 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?

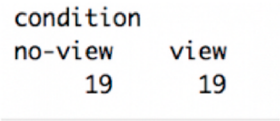
```
laptop_data$condition < - shuffle (laptop_data$condition)
```

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:

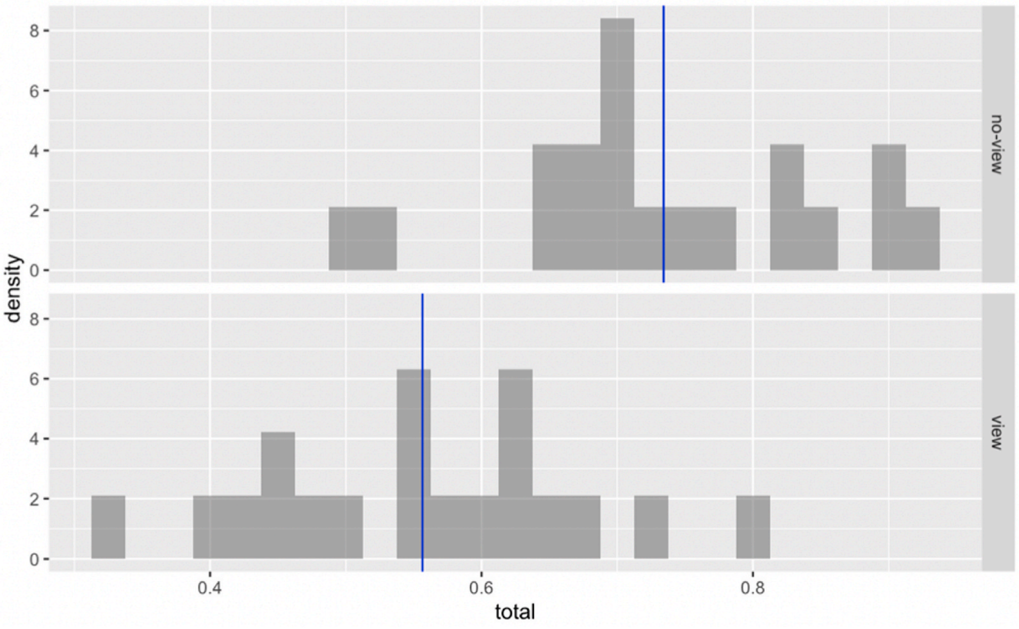
```
laptop_data$condition < - shuffle (laptop_data$condition)
tally (~condition, data = laptop_data)
```

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) ~ .)
```

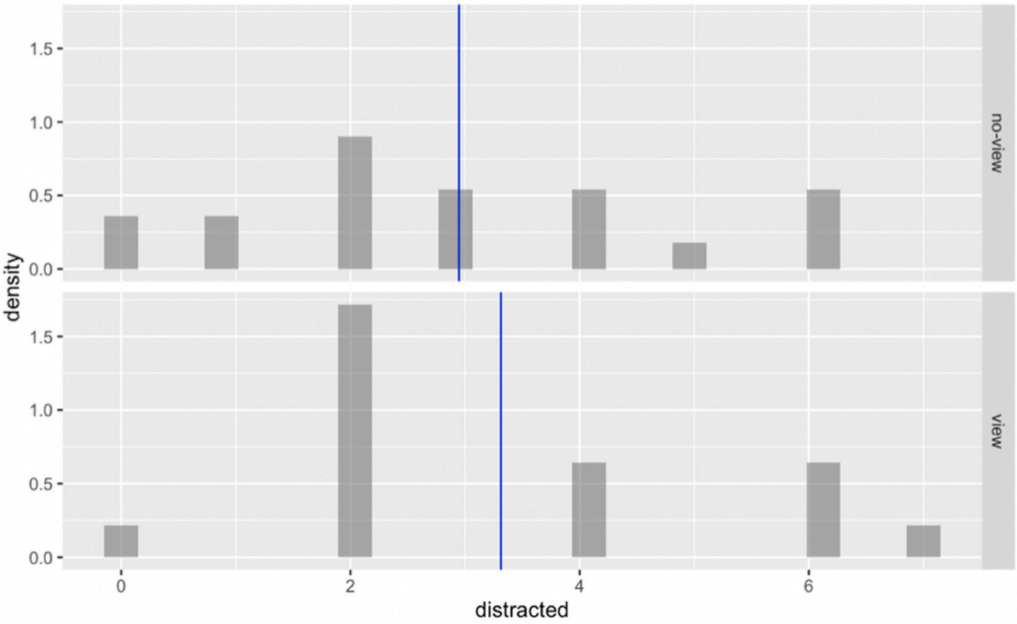
Example 2:

```
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**.

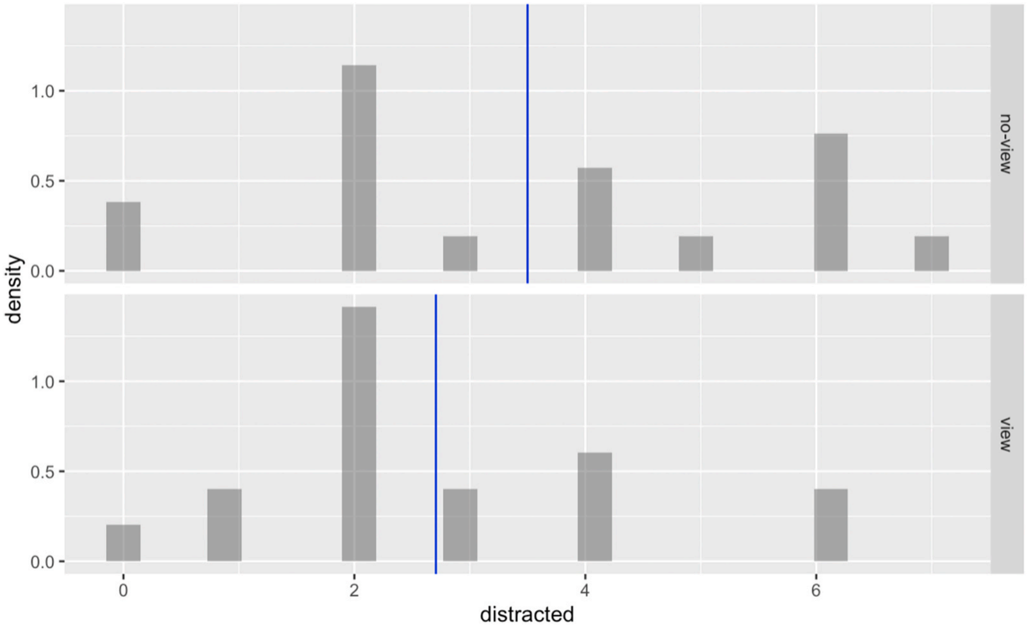
```
stats < - favstats (distracted ~ condition, data = laptop_data)
gf_dhistogram (~distracted, data = laptop_data) %>%
gf_vline (xintercept = ~mean, data = stats, color = “blue”) %>%
gf_facet_grid (condition ~ .)
```



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:

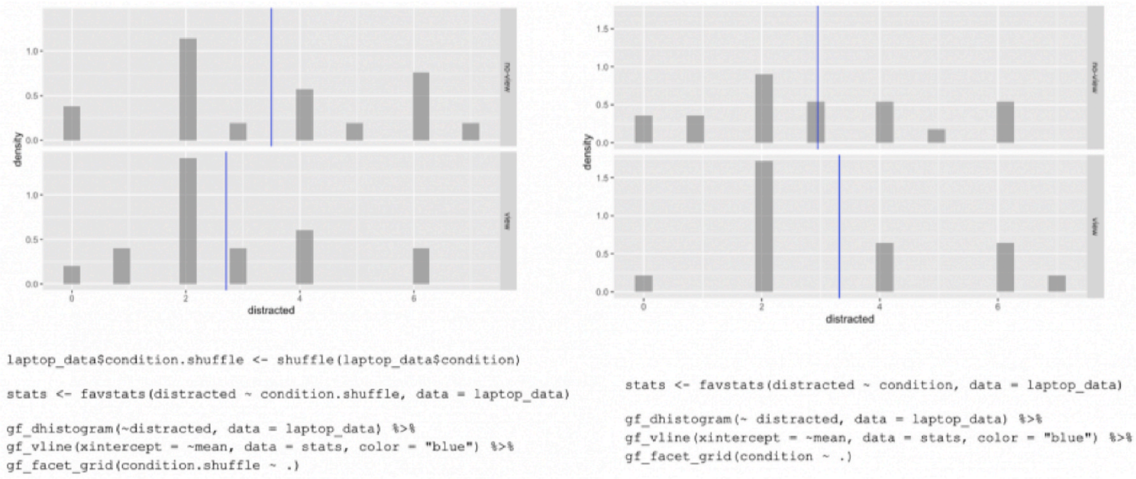
We revised the code from the previous question to create the graph below. We added a line to represent the mean of **distracted** for each **condition**. Notice that the average **distracted** rating in the **no-view condition** is higher than the average **distracted** rating in the **view condition**.

```
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):



```
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 ~ .)
```

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 1–6 (with 1 being not at all confident and 6 being extremely confident), how confident do you feel in your R skills?

References

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).

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. 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. https://doi.org/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). Dordrecht: Springer. 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. https://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

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* (p. 92). SERJ Editorial Board.

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. https://doi.org/10.1111/1467-9639.00033

Kastens, K. A., Liben, L. S., & Agrawal, S. (2008). Epistemic actions in science education. In *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*. 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., & Stigler, J. (2017-2022). *CourseKata Statistics and Data Science: A Modeling Approach*. Online. https://coursekata.org/.

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. https://doi.org/10.1598/0710.22, 441.

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 & 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, 13*(3). https://doi.org/10.1080/10691898.2005.11910562 [Online].

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.