Journal of Experimental Psychology: Applied

Representational-Mapping Strategies Improve Learning From an Online Statistics Textbook

Icy (Yunyi) Zhang, Maureen E. Gray, Alicia (Xiaoxuan) Cheng, Ji Y. Son, and James W. Stigler

Online First Publication, June 12, 2023. https://dx.doi.org/10.1037/xap0000474

CITATION

Zhang, I. (Y.), Gray, M. E., Cheng, A. (X.), Son, J. Y., & Stigler, J. W. (2023, June 12). Representational-Mapping Strategies Improve Learning From an Online Statistics Textbook. *Journal of Experimental Psychology: Applied*. Advance online publication. https://dx.doi.org/10.1037/xap0000474

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

https://doi.org/10.1037/xap0000474

Representational-Mapping Strategies Improve Learning From an Online Statistics Textbook

Icy (Yunyi) Zhang[1], Maureen E. Gray[2], Alicia (Xiaoxuan) Cheng[1], Ji Y. Son[3], and James W. Stigler[1]
[1] Department of Psychology, University of California, Los Angeles
[2] Department of Psychology, Elgin Community College
[3] Department of Psychology, California State University, Los Angeles

Using multiple representations is an important part of learning and problem-solving in science, technology, engineering and mathematics fields. For students to acquire flexible knowledge of representations, they must attend to the structural information within each representation and practice making relational connections between representations. Most studies so far have only attempted to help students connect between multiple representations in the lab or short-term classroom interventions, with the intervention largely separated from students' authentic learning. The present study developed a *representation-mapping intervention* designed to help students interpret, coordinate, and eventually translate across multiple representations. We integrated the intervention into an online textbook being used in a college course, allowing us to study its impact in a real course over an extended period of time. The findings of this study support the efficacy of the representation-mapping intervention for facilitating learning and shed light on how to implement and refine such interventions in authentic learning contexts.

Public Significance Statement

The study advances the idea that explicit representation-mapping can facilitate students' learning and transfer of statistics concepts. The findings provide important insights into college students' real learning behaviors and outcomes in an online environment. The method used in this study also guides the implementation of future theory-based interventions in authentic learning contexts.

Keywords: multiple representations, data visualizations, mapping, STEM, online learning

If we want our students to achieve flexible and transferable knowledge in science, technology, engineering and mathematics domains, we must help them develop *relational knowledge*. Knowledge of the relations and structures of a domain is what underlies experts' ability to transfer and flexibly adapt to novel situations (e.g., Ericsson et al., 2018). Relational knowledge connects and organizes the multitude of superficially different features, concepts, and representations that characterize complex domains (Goldwater & Schalk, 2016). For example, in math and statistics, relational knowledge allows students to connect variables (e.g., x and y) to other representations such as the coordinate plane (e.g., graphs) and equations, as well as other concepts such as mathematical operations and story contexts. Acquiring relational knowledge that supports adaptive, flexible thinking is typically more difficult than mastering a series of disconnected, isolated facts (Gentner & Kurtz, 2005). Eliciting transfer in a context novel to the student is, therefore, a notoriously difficult goal to achieve (e.g., Renkl et al., 1996).

Our focus here is on developing relational knowledge in statistics and data science. Once almost entirely based on mathematics

Icy (Yunyi) Zhang https://orcid.org/0000-0003-3423-6794
Maureen E. Gray https://orcid.org/0000-0003-4221-9185
Alicia (Xiaoxuan) Cheng https://orcid.org/0000-0001-8892-3683
Ji Y. Son https://orcid.org/0000-0002-4258-4791
James W. Stigler https://orcid.org/0000-0001-6107-7827

This project has been made possible in part by a grant from the Chan Zuckerberg Foundation and the California Governor's Office of Planning and Research Grant OPR18115. The authors have no conflicts of interest to disclose.

Icy (Yunyi) Zhang played lead role in data curation, formal analysis, investigation, project administration, visualization and writing–original draft and equal role in conceptualization, methodology and writing–review and editing. Maureen E. Gray played supporting role in investigation and equal role in conceptualization, methodology, resources and writing–review and editing. Alicia (Xiaoxuan) Cheng played lead role in validation and supporting role in conceptualization, data curation, methodology and writing–review and editing. Ji Y. Son played supporting role in conceptualization, methodology and supervision and equal role in funding acquisition, resources and writing–review and editing. James W. Stigler played lead role in supervision, supporting role in conceptualization, data curation and formal analysis and equal role in funding acquisition and writing–review and editing.

The study design, hypotheses, and analytic plans were not preregistered. The study data and analytic code can be accessed through https://osf.io/yxgm9/?view_only=dda142fd905248bf96d6312da921ed0f.

Correspondence concerning this article should be addressed to Icy (Yunyi) Zhang, Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095-1563, United States. Email: yunyi9847@g.ucla.edu

and mathematical models, statistics is increasingly becoming a computational science. Modern instruction extends beyond explanatory text and formulas to computational methods involving coding, data visualizations, and computer simulations. What comes with this trend is an increase in the number and types of representations necessary for students to master (Seufert, 2003). For example, Figure 1 shows a screenshot of a page in the online statistics textbook that is the setting for the current research (Son & Stigler, 2017–2023).

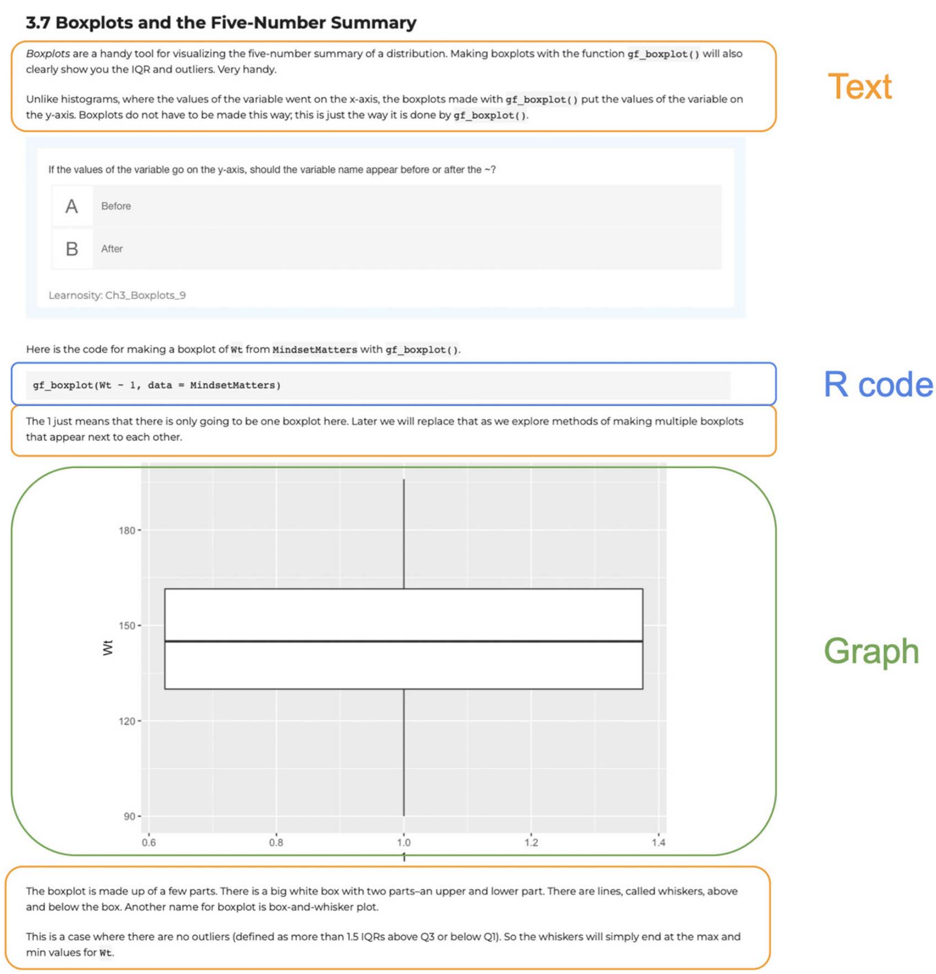
The goal of this page is to teach students about boxplots. But there are multiple related representations interleaved on the page—including text, computer code in the R programming language, and graphs—designed to help students better understand the boxplot. Research in math and statistics education has demonstrated the importance of connecting visual displays (such as graphs) with other representations, such as text or equations, in instructional materials (Renkl & Scheiter, 2017; Van Dooren et al., 2012). This makes modern statistics, with its proliferation of multiple representations, a rich and authentic domain for research on how connecting multiple representations can help students develop relational knowledge (Star & Rittle-Johnson, 2009).

Working with multiple representations is an important part of expert thinking in a domain (Larkin & Simon, 1987). For novices to become experts, they need to develop the ability to translate and connect across multiple representations (i.e., representational flexibility; for a review, see Acevedo Nistal et al., 2009; or within- and between-representation fluency, for a review, see Star & Rittle-Johnson, 2009). Representational flexibility grows as learners recognize the structural relationships *within each representation* and leverage these relationships to make connections *across multiple representations* (Even, 1998; Lin et al., 2016; Rau et al., 2015; Seufert, 2003; Waisman et al., 2014). These connections within- and between-representations are fundamental to understanding and transfer in a domain (Chang et al., 2016; Star & Rittle-Johnson, 2009).

Research suggests that exposure to multiple representations of the same concept can benefit students' learning in science, technology, engineering and mathematics domains (Acevedo Nistal et al., 2009; Cheng, 2000; Mayer, 2009; van der Meij & de Jong, 2006). Working

Figure 1

Screenshot of Multiple Representations in an Online Statistics Textbook (IQR in the Screenshot Stands for Interquartile Range)



Note. See the online article for the color version of this figure.

with multiple representations can make knowledge more adaptive and flexible (Spiro, 1988), while compensating for the limitations of each representation (de Jong et al., 1998; Gagatsis et al., 2006). Activities in which students connect multiple representations for the same concept—for example, the equation and graph for a straight line in the context of linear regression—can support the development of adaptive expertise (de Jong et al., 1998; Kellman et al., 2010; McGee & Moore-Russo, 2015).

Despite these benefits, getting students to engage in the work required to form interconnected knowledge of representations can be a challenge (Ainsworth et al., 2002; Bodemer et al., 2004; Even, 1998; Lin et al., 2016; Mason et al., 2013; Renkl & Scheiter, 2017; Schwonke et al., 2009; Waisman et al., 2014). Past research has shown how difficult it is for students to interpret and operate flexibly within and between multiple representations (Star & Rittle-Johnson, 2009). Students struggle to navigate multiple representations and often do not make effective use of them (Ainsworth et al., 2002; Lin et al., 2016; Van Dooren et al., 2012; Van Someren et al., 1998). Further, mere exposure to multiple representations is not sufficient for developing representational flexibility (Ainsworth, 2006; Rau et al., 2015). Representations, such as visual displays, graphs, or equations, each impose a high cognitive demand on learners, which may be compounded when learners are shown multiple representations (Renkl & Scheiter, 2017; Van Dooren et al., 2012).

Eye-tracking evidence also demonstrates the tensions and difficulties around comprehending multiple representations, compared to a single representation (e.g., text alone). Eye-tracking studies have found that when shown both text and visualizations together on a page, learners pay attention to the text at the expense of comprehending the visualization (Hannus & Hyönä, 1999; Hegarty & Just, 1993; Schmidt-Weigand et al., 2010). When the illustrations are irrelevant to the text, these images can distract learners (Sanchez & Wiley, 2006). Even when the images are relevant to the text, coupling illustrations with text can lead to overconfidence and an illusion of learning compared to seeing text alone (Ackerman & Leiser, 2014; Jaeger & Wiley, 2014; Wiley, 2019).

Part of the difficulty in connecting multiple representations is that interpreting each representation requires some background knowledge, which the learner may or may not possess (Kozma & Russell, 1997; Renkl & Scheiter, 2017). Novices may not make appropriate connections among the various representational formats because the correspondences that seem intuitive to domain experts are not obvious to novice learners (e.g., Chi et al., 1981). For example, an eye-tracking study conducted by Mason et al. (2013) showed that when reading, learners with greater prior domain knowledge were better able to integrate texts and visual displays than were learners with less domain knowledge. Novices can also be more easily distracted by irrelevant features in a visualization (Hegarty et al., 2010; Lowe, 2004). When learners just begin to learn about a new representation (e.g., a set of boxplots), they do not already know which features are informative (the relative position of the boxes) and which are not (whether the ends of the whiskers have perpendicular caps on them or not).

Researchers and educators have explored various strategies to help novices connect multiple representations. Some studies have focused on physical integration, for example, putting text into a graph instead of separating the two (for a review, see Ayres, 2020). Others have sought to leverage multiple modalities, such as using spoken text simultaneously with dynamic visualizations (for a review, see Low & Sweller, 2014) or using contrast cases to help students find deep structures underlying the multiple representations (Schwartz et al., 2011). Our focus here is on strategies that directly highlight the underlying relationships within- and between-representations, such as verbal explanations and gestures that draw attention to relational connections between multiple representations (Berthold & Renkl, 2009; Richland, 2015; Seufert, 2003; Seufert & Brünken, 2006).

Pedagogical approaches such as these are reminiscent of the analogical learning literature. Analogical learning strategies, such as alignment, mapping, and comparison, are often used to increase conceptual understanding and transfer because they require students to attend to relational structure (Alfieri et al., 2013; Fries et al., 2021; Gentner et al., 2003; Gentner & Maravilla, 2018; Goldwater & Schalk, 2016; Gray & Holyoak, 2021; Holyoak, 2012). These strategies can be directly applied to the goal of making connections between multiple representations. For example, aligning two representations can help students connect superficially dissimilar features that have the same underlying function (e.g., the *x*-axis in a histogram represents the possible values of a variable, whereas the *y*-axis serves that function in a vertical boxplot).

Past research has shown the benefits and importance of highlighting alignment and comparisons between structurally similar concepts across different representations (e.g., Butcher, 2006; Martin et al., 2019; Scheiter & Eitel, 2015; Schmidt-Weigand et al., 2010; Thompson & Opfer, 2010). For example, Scheiter and Eitel (2015) discovered that students who were taught with visual cues emphasizing the relationship between visual diagrams and text learned more effectively than pedagogy without such cues. Similarly, integrating and making explicit the connections between multiple representations such as diagrams and numerical representations supports learning (van der Meij & de Jong, 2006).

However, the few studies that have demonstrated such effects of highlighting alignment and comparisons between multiple representations were either conducted in the laboratory (e.g., Schmidt-Weigand et al., 2010) or as very short-term classroom interventions (e.g., Martin et al., 2019), using content that was not integrated into the normal class learning materials (e.g., Berthold & Renkl, 2009; Seufert, 2003). Moreover, most multiple representation research has mostly focused on just two types of representations (e.g., graphs of lines and equations), though in most domains, students often must navigate among a rich interconnected network of representations (e.g., boxplots, histograms, code, equations, verbal descriptions).

The Present Study

Our goal in the present study is to use insights from cognitive psychology and the multiple representations literature to design, integrate, and evaluate an intervention to help students make connections and extract schemas across multiple representations in an authentic online learning context. The context in which the study was conducted was a semester-long college-level introductory statistics course.

We started by identifying a few representations that need to be learned *and* connected in students' minds in order to make their statistical knowledge more coherent. Our instructional goal was to deepen students' understanding of histograms and boxplots by helping them to connect these graphical representations to R code, word equations, and verbal descriptions of situations in the world. In our experience, it takes time for students to learn about these two different types of data visualizations. Even when the same data are being visualized with both a histogram and a boxplot, students may not attend to the structural similarities and differences between the two types of graphs and the R codes that generate them.

Simply teaching students about different representations does not guarantee that they will be able to connect them at more than a superficial level. Achieving representational flexibility with these data visualizations requires students to understand the connections between the visuo-spatial information in a plot and the underlying semantic information represented by the plot (Schnotz & Bannert, 2003).

A representation-mapping intervention that uses alignment and comparison of histograms, boxplots, and the code used to produce them, can facilitate this process by helping learners make connections between the relational structures the different types of plots have in common. Making these connections explicit by drawing lines and annotating graphs may be an especially helpful strategy in multimedia learning (Jee et al., 2013; Thompson & Opfer, 2010). Presenting aligned examples in a visually salient manner facilitates comparison and structural learning (Jee et al., 2013; Thompson & Opfer, 2010). Beyond visual presentation, students are aided when instructors explicitly emphasize the connections between structures in class (Richland et al., 2007).

To design our mapping intervention, we identified key connections among four types of representations: the plots that were the focus of the instruction (i.e., histograms and boxplots); the R code used for creating the plots, which would help to highlight the structure underlying the plots; verbal descriptions of the plotted distributions; and word equations that represent the relationships in graphs. The goal is to help students form a more relational, interconnected, and thus coherent understanding of the statistical concepts and also learn how to actively engage in such comparisons on their own.

A secondary goal of the study was to explore a new approach for moving learning strategies from the laboratory into authentic learning settings. Toward this end, we employed the better book approach to education R&D (Stigler et al., 2020) in which experiments are codesigned with researchers, instructors, and developers and implemented in the context of an online interactive textbook being used in real courses. Students participating in this study were enrolled in college statistics courses using the textbook, *Statistics and Data Science: A Modeling Approach* (Son & Stigler, 2017–2023).

The online textbook was hosted on the CourseKata platform (https://coursekata.org), which enables researchers to randomly assign students within the same class to slightly different versions of the material (Stigler et al., 2020). In the present study, students were randomly assigned to a version of the textbook that either included the representational-mapping intervention or not. A de-identified version of the data generated by students as they work through the book was made immediately available to researchers on the CourseKata platform.

Working in an authentic course environment yields several advantages. First, whatever we find in our research can result in immediate improvements to the instructional materials. If we find that our representational-mapping intervention improves students' learning in this course, such a finding would not only appear in a scholarly publication but also would be stored in the form of improvements to the online textbook, to be used by future instructors and students.

Another advantage of working in the context of an authentic course is that it gives us a more detailed understanding of students' knowledge prior to the intervention. In the present study, for example, we knew that because students were several weeks into the course at the time of the intervention, they had already learned some basic R concepts and procedures, understood the structure of a data frame in R, and knew the difference between a variable and an observation. This made it possible to design an intervention that fit with students' prior knowledge. And because this experiment was implemented as part of normal homework for a real course, students had time to learn these concepts at a more natural pace.

The online, interactive textbook (Son & Stigler, 2017–2023) was made up of pages that interleaved text, videos, questions, and R coding exercises. The textbook was designed according to the practicing connections approach to curriculum design (Fries et al., 2021). The approach, simply stated, is this: if we want students to develop coherent and flexible knowledge in complex domains, they must be provided with opportunities to practice making connections—among core concepts, key representations, and a broad range of relevant contexts—and not just the isolated skills and concepts required for routine expertise in the domain.

The book takes a modeling approach to statistics, connecting the typical concepts and skills taught in introductory statistics to the practice and conceptual structure of statistical modeling (see Son et al., 2021, for a comprehensive overview of the curriculum). It emphasizes several key representations, including graphs, R code, word equations, and notation of the general linear model to represent relationships among variables, and it is highly interactive, including more than 1,200 formative assessment questions and R coding exercises.

We developed our mapping intervention as a set of supplementary videos that could be embedded in the book at the points where students are learning about histograms and boxplots (Chapters 3 and 4). Links to the five intervention videos are included in Appendix A. The videos were designed to direct students' attention to the structural similarities among multiple representations by mapping the same information across representations. Half the students taking a college-level introductory statistics course were randomly assigned to a version that included the embedded videos, while the other half saw a version of the book that did not include the videos. The question of interest was whether students' engagement with the videos would enhance their learning of important structures and relations in the domain.

Because the book is already designed to support students' practicing of connections, it provides an even more rigorous testing ground for our specific hypotheses about this representation-mapping intervention. We hypothesized that controlling for students' performance in the earlier chapters of the book, students who watched the intervention videos would outperform students who did not watch the videos. We also hypothesized that the effects of watching the videos on learning would be mediated by students' own use of the representational-mapping strategies illustrated in the videos.

Method

Participants

Participants were 210 undergraduate students taking an introductory statistics course (*Introduction to Psychological Statistics*) that used the aforementioned interactive textbook at University of California, Los Angeles. They participated in the study as part of their regular homework assignments, which involved reading and responding to questions in the online textbook. 104 students were randomly assigned to the experimental condition (with two pages that included the supplementary videos), and 106 to the control condition (which included the same pages but without the videos).

Twenty-four additional students in the class were excluded from analyses either because they did not complete the course and receive a final grade (21 students), or because they scored 0 on all of the outcome measures, indicating a lack of effort and engagement (three students).

Although our sample size was predetermined by the number of students in the class, we still wanted to know if it was large enough to detect, at the very least, a small to medium effects. Toward this end, a power analysis was conducted using the pwr package in R (Champely et al., 2017; Cohen, 1988). The analysis indicated that we would be able to detect a small to medium effect of condition ($f > .05$), and at least a medium effect of the number of videos watched ($f > .10$), on the outcome measures.

The study was reviewed and approved by the university's institutional review board for the protection of human participants.

Representation-Mapping Intervention

The representation-mapping intervention took the form of five instructional videos, embedded at two points in the online textbook (in Chapters 3 and 4). The online textbook contains 12 chapters broken into three parts: exploring variation, modeling variation, and evaluating models. Chapters 3 and 4, where the intervention and in-book assessment took place, were in the exploring variation part of the textbook. Chapter 3 introduced the concept of a distribution and used different types of graphs to visualize different types of variables (e.g., histograms, boxplots, and bar graphs). Chapter 4 introduced the concept of using one variable to explain variation in another (e.g., we might use variation in students' sex to explain variation in their thumb lengths). Students learn how to graph relationships between two variables, look for evidence in the graphs that one variable explains variation in another, and represent these relationships with informal word equations (e.g., thumb length = sex + other stuff).

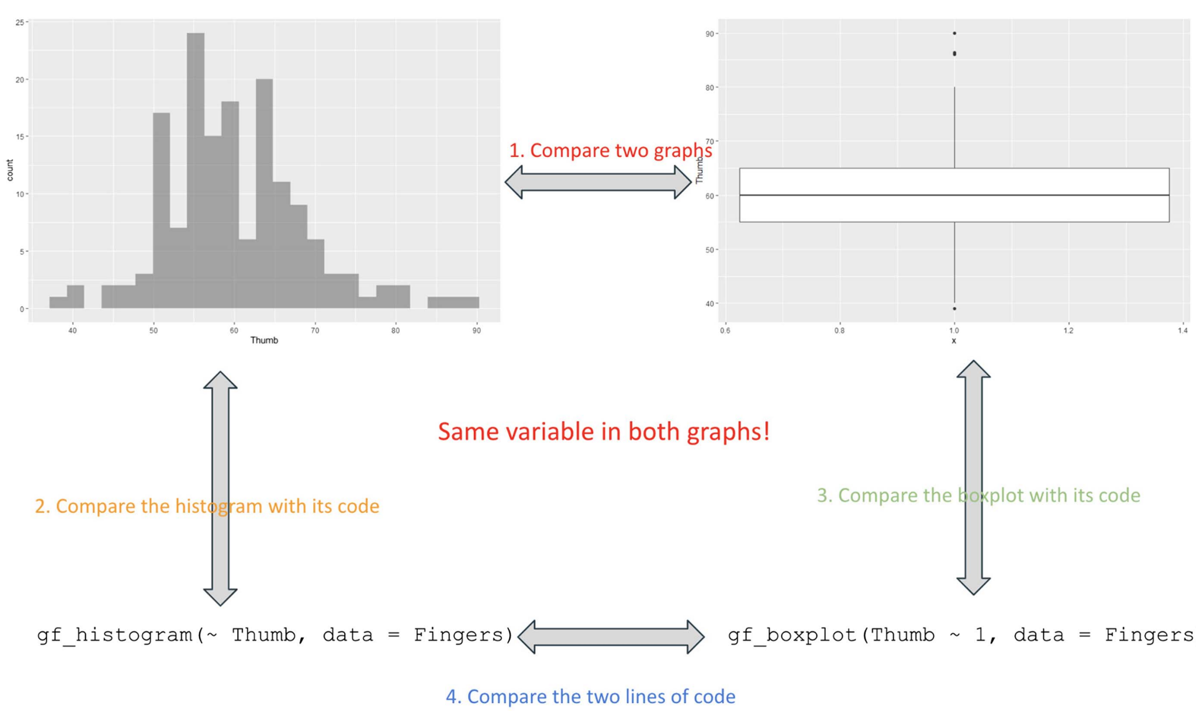
The five videos that comprise the intervention run for a total of 26 min. (The first video was 9 min and 11 s; the second video was 3 min and 26 s; the third video was 3 min and 33 s; the fourth video was 6 min and 6 s; the fifth video was 3 min and 43 s.) The videos were designed to highlight shared structure among different representations of the same statistical relationship.

In each video, an instructor explicitly pointed out the alignment between four different types of representations: text, R code, data visualizations, and informal "word equations." We identified this particular set of representations because (a) research has shown that students have difficulty interpreting and translating between different forms of numerical, graphical, symbolic, and verbal representations (Bossé et al., 2011; Van Dooren et al., 2012) and (b) instructors who have used these course materials anecdotally report that students struggle to connect visualizations with corresponding R code at this point in the course.

Figure 2, for example, shows the four comparisons highlighted in one of the intervention videos in which the instructor explicitly discusses the alignment among histograms, boxplots, and the R code used to generate them. Although both graphs depict the same

Figure 2

An Example of Mapping Between Multiple Representations in the Intervention



Note. See the online article for the color version of this figure.

distribution (the distribution of students' thumb lengths measured in mm), the superficial differences between them make it difficult for students to connect and align the two types of plots. For example, while histograms represent the values of a quantitative variable on the *x*-axis, boxplots represent those same values on the *y*-axis. This distinction is represented in R by the placement of the tilde. Putting the variable name before the tilde (e.g., Thumb ∼) results in the variable being represented on the *y*-axis, whereas putting the variable after the tilde (e.g., ∼ Thumb) results in it being represented on the *x*-axis.

Each of the five intervention videos consisted of a screen recording of a female instructor (appearing in a small inset window) discussing graphs and R code as they appeared on a series of slides. The instructor drew annotations on each slide as she discussed it, all of which was captured on the screen recording. By using lines and colors to explicitly map similar components between multiple representations, each video was designed to focus learners' attention on key connections.

The first three videos were embedded on Page 7 in Chapter 3 (Page 3.7) of the book (see Figure 3). In the first and second videos, the instructor compared and contrasted two visualizations (a boxplot and a histogram) of the distribution of a single variable based on the same data (depicted by arrow 1 in Figure 2), and showed how the five-number summary (i.e., the maximum, minimum, lower and upper quartiles, and median) can be mapped to each plot. In the third video, the instructor discussed the R code that generated each graph and connected features of the graphs with features of the R code (depicted by arrows 2, 3, and 4 in Figure 2).

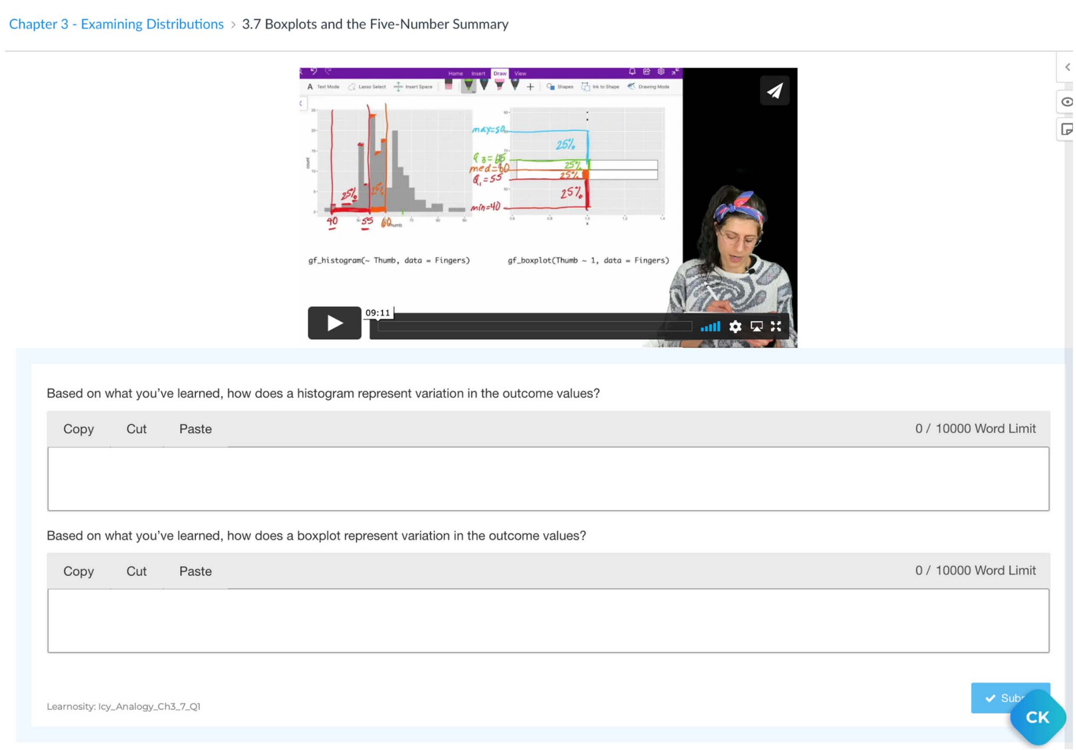
The final two intervention videos appeared on Page 5 in Chapter 4 (Page 4.5). The fourth video again compared histograms, boxplots, and R code but this time in the context of bivariate relationships (instead of univariate distributions). For example, Figure 4 shows side-by-side boxplots, with one variable, Thumb, on the *y*-axis, and another variable, Sex, on the *x*-axis. The core underlying connection between the R code, boxplots, and verbally described hypothesis is that all three represent the idea that Sex predicts some of the variation in Thumb. The fifth video, focused on connections between the R code (e.g., gf_boxplot(Thumb ∼ Sex)) and an informal word equation used to represent a relationship between two variables (Thumb = Sex + Other Stuff).

The mapping intervention videos follow a developmental trajectory. Learners begin in Chapter 3 by connecting multiple representations (histograms, boxplots, and the five-number summary) of a single variable distribution (e.g., Thumb lengths). Later in Chapter 3, they connect these representations to R code and word equations. In Chapter 4, students extend their understanding of these visual representations of single variables to explore relationships between two variables (e.g., Thumb length and Sex).

For example, students initially learn that the *y*-axis on a vertical boxplot represents values of the variable Thumb but that the *x*-axis is meaningless. But then in Chapter 4, students learn about side-by-side boxplots where the *x*-axis is now used to represent values of the second, categorical variable, Sex (e.g., female or male). This knowledge about axes connects to the R code for producing the boxplots, where the variables on the *y*- and *x*-axis must be entered in a particular order (Y ∼ X, e.g., gf_boxplot(Thumb ∼ Sex)). Having

Figure 3

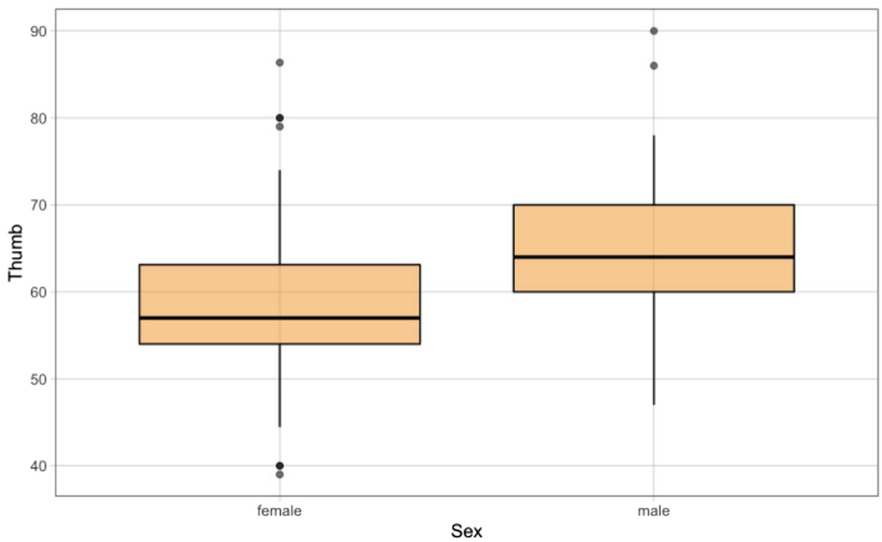
Screenshot of the Experimental Page With an Intervention Video From Chapter 3



Note. See the online article for the color version of this figure.

Figure 4
R Code and Side-by-Side Boxplots of Bivariate Relationships

```
gf_boxplot(Thumb ~ Sex, data=Fingers)
```



Note. See the online article for the color version of this figure.

well-mapped connections between these multiple representations could potentially facilitate students' learning later in the book when they will create linear models using a similar syntax.

It is worth noting that the concepts and skills referenced in the intervention videos had already been covered in the textbook. For example, on prior pages students had already been exposed to histograms, including explanations of how to interpret them and create them in R. The goal of the intervention videos, therefore, was *not* to introduce new information but instead to make explicit connections among different representations of the same statistical ideas.

Procedure

Students taking the class were randomly assigned to get one of the two versions of the textbook, which they were expected to complete for homework. The only pages that differed between the two versions were Page 3.7 (in Chapter 3) and Page 4.5 (in Chapter 4). Students in the experimental condition saw versions of these pages that included the intervention videos (three on Page 3.7, two on 4.5) while students in the control condition saw versions that did not include the videos. Learning from the intervention videos was assessed using questions embedded throughout Chapters 3 and 4 of the book as well as with instructor provided measures, that is, quiz scores.

Measures

Pre-Intervention Assessment

Students' performance in the course prior to the start of the intervention (in Chapter 3), was measured using 32 multiple-choice review questions assigned at the end of Chapter 2. Students received class credit just for attempting these questions but for research purposes, their responses were scored as correct or incorrect and then summed to get a total number correct.

In-Book Assessments

Students' learning was primarily assessed by a subset of the many questions embedded throughout Chapters 3 and 4 of the interactive textbook. These in-book assessments were a mix of multiple-choice and free-response questions designed to assess students' understanding of histograms, boxplots, and the alignment between the two. Because the questions were asked to both experimental and control students, they did not make reference to the contents of the intervention videos; all students could, in theory, have answered the questions correctly just based on information available to everyone in the textbook. For example, we asked, "Based on what you've learned, how does a histogram represent variation in the outcome variable?" and, "If we made a histogram with four bins (bins = 4), would those bins cut the data at the same points as the five-number summary?"

Figure 5 shows where the in-book assessment questions were placed in Chapters 3 and 4. After the Chapter 3 intervention videos (on Page 3.7), there were 30 in-book questions (7 multiple-choice, 23 free-response) relevant to the content of the intervention that appeared on Pages 3.7 through 4.5. We will refer to these as Chapter 3 in-book questions because the content was related to the Chapter 3 intervention videos. After the Chapter 4 intervention videos (on Page 4.5), we identified 19 questions as Chapter 4 in-book questions (9 multiple-choice, 10 free-response). These questions were spread out across the rest of the pages in Chapter 4.

All in-book questions were posed to both experimental and control groups. Students were given one point for each correct response. We also collected data from the online platform to get a

Figure 5
Placement of Questions in the Book

	Chapter 3					Chapter 4									
Pages	3.1 … 3.6	3.7	3.8	3.9	3.10	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	4.10
Number of intervention videos		2								3					
Number of questions		"Chapter 3 questions" 30 questions (7 multiple-choice and 23 free-response)								"Chapter 4 questions" 19 questions (9 multiple-choice and 10 free-response)					

Note. See the online article for the color version of this figure.

rough measure of how much time students spent studying Chapters 3 and 4 of the textbook.

Delayed Take-Home Quizzes

We used two take-home quizzes that were administered by the instructor to assess students' performance after a delay. We will refer to the first quiz as the *near-transfer* quiz, and the second as the *far-transfer* quiz.

The near-transfer quiz was administered 1 day after students were required to have completed their work in Chapters 3 and 4 of the textbook. The quiz questions were similar to those asked in the textbook but used a different context and a different data set. Students were asked on the near-transfer quiz to write and/or run R code to create appropriate visualizations for variables in a new data set and then to interpret those visualizations, a task which was similar to what they had done in the textbook. For example, one question asked students to run some code to create a histogram with 10 bins and then asked them to explain why some of the bins appeared to be empty. The near-transfer quiz consisted of 18 questions (7 multiple-choice, 11 free-response). Students were given 10 hr to complete the quiz. The quiz was scored on a scale of 0–100.

The far-transfer quiz was administered near the end of the quarter (around 6 weeks after the intervention), and the questions were less similar and more challenging than those in Chapters 3 and 4. The far-transfer quiz contained 16 questions (11 multiple-choice, 5 free-response). It evaluated students' understanding of the entire course, including visualization related concepts covered in Chapters 3 and 4 and required students to engage in statistical inferences using a real data set. Chapter 4 concepts regarding the relationships between variables were fundamental to the development of formal statistical models later in the course. Students' understanding of formal models was the main learning outcome assessed in the far-transfer quiz. The quiz was also scored on a scale of 0–100.

Video-Watching Behavior

Just because the intervention videos were embedded in the textbook did not mean that students would necessarily watch them. Thus, we measured the number of intervention videos each student watched on each of the two pages. Because there were also other nonintervention videos in the textbook, we also measured the number of nonintervention videos students watched in Chapter 4, as a control variable (there were no nonintervention videos in Chapter 3).

For all videos, we defined "watched a video" as having watched a proportion of .5 or more of the video. We chose to use this binary measure of students' video-watching behavior because the distribution of students' proportion of each video watched was highly bimodal, with video watching proportions either near 0 or 1. For example, for the first intervention video in Chapter 3, 89.2% of the proportions were either below .05 or above .95, and 95.7% for the second intervention video.

Use of Compare and Contrast

Finally, in addition to coding questions for correctness in order to measure students' learning, before data collection commenced we selected five free-response questions from the Chapter 3 in-book assessments (i.e., Page 3.7–4.4) to be coded for explicit use of compare and contrast strategies. These questions were chosen either because they directly prompted students to engage in comparison (e.g., "Compare and contrast histograms with boxplots. What are the strengths and weaknesses of each?") or because, in the opinion of the researchers, the quality of students' answers would be improved if they employed compare and contrast strategies.

Students' responses on each of these questions were coded as 0 or 1, with 1 indicating that they used compare and contrast strategies to answer the question. We defined use of compare and contrast strategies by the inclusion of comparative words (e.g., whereas, compared to, more, less) to describe the relationship between two concepts (e.g., boxplot vs. histogram; quantitative variable vs. qualitative variable). Students' total scores for use of compare and contrast strategies thus could range from 0 to 5.

Transparency and Openness

Data

The data are published on the Open Science Framework website and the link to access this information is provided in https://osf.io/yxgm9/?view_only=dda142fd905248bf96d6312da921ed0f.

Analytic Methods

The R code needed to reproduce analyses is available on the Open Science Framework website and the link to access this information is provided in https://osf.io/yxgm9/?view_only=dda142fd905248bf96d6312da921ed0f.

Materials

The links to the intervention videos and the in-book assessment questions are included in Appendices A and B.

Data Analysis

Coding

Five trained coders coded students' responses to free response questions both for correctness and for the use of compare and/or contrast strategies. Each coder was randomly assigned to code a subset of students' responses blind to condition and each response was coded independently by two coders. For some of the free-response questions, partial credit scores of 0.5 were given.

Three questions for which the two coders disagreed on more than 25% of responses were discussed with the whole team. After the discussion, the scoring rubric for those questions was revised and all responses recoded. The discrepancy rate averaged across all coders and questions was 10.9%. These discrepancies were resolved in meetings that included the two coders and at least one additional coder. Cohen's κ statistic was calculated to assess interrater reliability of the coding. The agreement between raters was substantial, $\hat{e} = .7$, and greater than would be expected by chance, $p < .001$.

Planned Analysis

For each learning assessment, we were not only interested in the difference between experimental groups (intervention vs. control) but also in the relationship between watching videos and performance on the assessment.

Our measure of the number of videos watched differed depending on which outcome assessment we wanted to analyze. If we were analyzing the in-book assessments in Chapter 3, we used only those intervention videos that students had watched in Chapter 3. For assessments in Chapter 4 and beyond, we used the total number of intervention videos watched in both Chapters 3 and 4. Our assumption was that video watching could affect only those assessments that students completed after they had watched the video.

We also evaluated the impact of watching nonintervention videos by including it as a covariate when analyzing the near- and far-transfer assessments. This helps to rule out selection effects related to video watching and makes clear the unique contribution of watching intervention videos.

Moreover, we planned to use median analyses to explore whether students' improvement in near- and far-transfer assessments can be explained by their use of comparing and contrasting strategies in the book assessments.

Results

Nonnormality of Performance

All measures of student engagement and performance (e.g., prior performance, in-book performance, near- and far-transfer, use of compare and contrast strategies) showed various degrees of nonnormality. Shapiro–Wilk tests confirmed that all distributions departed significantly from normality, ps < .001. (For a detailed summary of the nonnormality test statistics, see Appendix C.)

Because of nonnormality in the dependent variables, we relied on two nonparametric statistical tests: the Wilcoxon-rank sum test and bootstrapped confidence intervals (CI). To examine any effect of condition, we performed Wilcoxon rank-sum tests.

Whenever we wanted to control for other variables (e.g., prior performance), we created 10,000 bootstrapped samples from the data. For each sample, we fitted a model that included all variables including those we wanted to control for (e.g., preintervention performance). We then calculated for each sample the median difference between groups or the slope representing the relationship between two quantitative variables (e.g., number of videos watched and performance). We used these bootstrapped estimates to create a 95% confidence interval. For all bootstrapped analyses, we used the boot package in R (v1.3–27; Canty & Ripley, 2021).

Preintervention Performance

We analyzed students' performance on the review questions at the end of Chapter 2 in order to see if the experimental and control groups differed before the onset of the intervention. Scores ranged from 0 to 32 correct among students in the control group ($Mdn = 28$) and from 8 to 32 for the experimental group ($Mdn = 28$). A Wilcoxon rank-sum test showed no significant difference between the medians of the two groups ($Z = -0.31$, $p = .753$, $r = .02$).

Video-Watching Behavior

The number of intervention videos and nonintervention videos students in the experimental and control groups watched is shown in Table 1. On the whole, students in the experimental group did not tend to watch the intervention videos. But then they also did not tend to watch the nonintervention videos, a pattern they shared with students in the control group. The average number of intervention videos watched by students in the experimental group, of the three in Chapter 3, was 1.30, and of the two in Chapter 4, 0.83. The average number of nonintervention videos watched of the two in Chapter 4 was 0.85 for the experimental group, and .88 for the control group. These averages did not differ significantly between the two groups, $t(208) = 0.35$, $p = .727$. 88.5% of students in the experimental group watched at least one of the videos in Chapter 3; only 71.2% watched at least one intervention video in Chapter 4.

Time Spent on Chapters 3 and 4

Students spent a total of 408.65 min on average studying Chapters 3 and 4 ($SD = 311.58$ min); the median time spent was 344.20 min. Figure 6 shows the distribution of students' total time spent in Chapters 3 and 4 broken down by condition. A Wilcoxon rank-sum test showed that the experimental group ($Mdn = 382.49$) and control group ($Mdn = 301.22$) did not differ significantly on the time they spent on the two chapters ($Z = -1.63$, $p = .103$, $r = .12$). A bootstrap analysis showed that, within the experimental condition, neither the number of intervention videos watched (95% bootstrapped CI [−28.23, 95.37]) nor the number of nonintervention videos watched (95% bootstrapped CI [−77.96, 165.38]) significantly predicted the total amount of time each student spent on Chapter 3 and 4. Within the control group, the number of nonintervention videos watched

Table 1
Distribution of Students in Experimental and Control Group by Number of Intervention and Nonintervention Videos Watched in Chapters 3 and 4

	Experimental group			
Students who watched …	Chapter 3 intervention videos	Chapter 4 intervention videos	Chapter 4 nonintervention videos	Control group Chapter 4 nonintervention videos
0 videos	12	30	27	33
1 video	55	62	66	53
2 videos	31	12	11	20
3 videos	6	NA	NA	NA
At least 1 video	92 (88.5%)	74 (71.2%)	74 (74.0%)	73 (68.9%)

Note. NA = not applicable.

also did not significantly predict their time spent in these two chapters (95% bootstrapped CI [−20.37, 145.03]).

In-Book Performance

Chapter 3

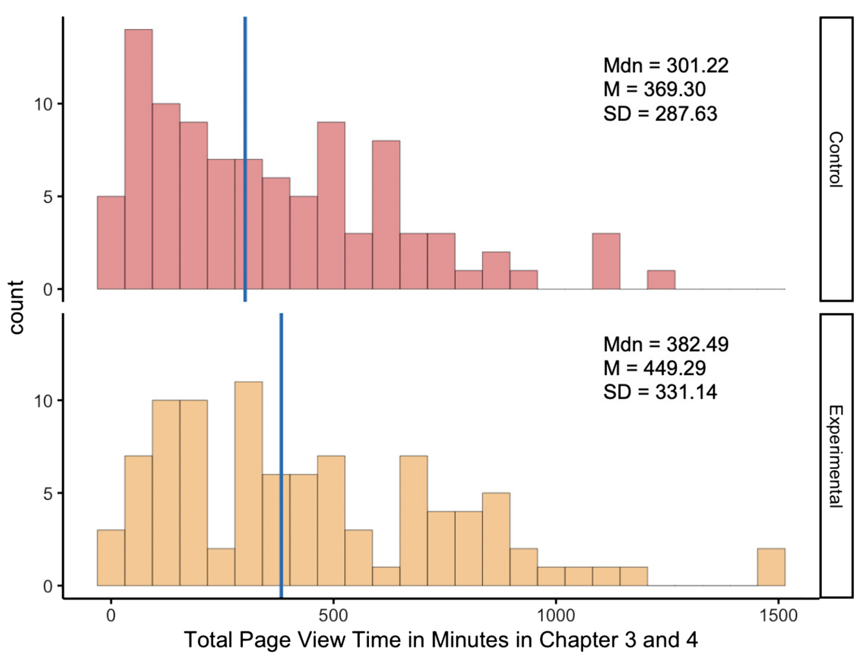
Summary statistics for students' performance in Chapter 3, broken down by condition, are reported in Table 2. A Wilcoxon rank-sum test showed that students in the experimental condition ($Mdn = 21$, $n = 104$) performed significantly better than students in the control condition ($Mdn = 20$, $n = 106$; $Z = 2.69$, $p = .007$, $r = .19$). A bootstrapped confidence interval based on 10,000 bootstrapped samples and controlling for preintervention performance showed the same effect (95% bootstrapped CI [0.50, 2.67]). Full tables of the analyses were included in Appendix E.

Because not all students in the experimental group watched the Chapter 3 intervention videos, we also examined whether the number of Chapter 3 intervention videos watched predicted students' performance for Chapter 3 in-book assessments within the experimental group (Figure 7, upper left; Table 3). Controlling for students' prior performance in the book, the 95% bootstrapped confidence interval for the slope of number of videos watched predicting Chapter 3 in-book assessments was [0.12, 1.98], an interval that did not include 0.

Chapter 4

Unlike Chapter 3, the Wilcoxon rank-sum test showed no significant difference between the control group ($Mdn = 13$, $n = 106$) and the experimental group ($Mdn = 12$, $n = 104$), $Z = .33$, $r = .02$, $p = .739$ for performance on the Chapter 4 in-book assessments. Even when controlling for prior performance, the difference between the two groups was not significantly different from 0 (bootstrapped 95% CI [−0.98, 0.98]).

Figure 6
Time Spent on Chapters 3 and 4 by Condition (in Minutes)



Note. See the online article for the color version of this figure.

Table 2
Chapter 3 Performance by Condition

Condition	Chapter 3 in-book performance			Chapter 4 in-book performance		
	Mdn	*M*	*SD*	*Mdn*	*M*	*SD*
Control	20	19.04	4.99	13	11.25	4.00
Experimental	21	20.70	4.65	12	11.35	4.04

One reason we did not find an effect in Chapter 4 might be that even fewer students watched the intervention videos than in Chapter 3 (88.5% watched at least one video in Chapter 3 vs. only 71.2% in Chapter 4). We conducted an additional analysis on participants who had access to the intervention videos to examine whether the cumulative number of intervention videos watched in Chapter 3 and 4 predicted performance on Chapter 4 in-book assessments (Figure 7, upper right; Table 4). Controlling for students' prior performance in the book, we found a significant relationship between the total number of intervention videos watched in Chapter 3 and 4 and students' performance on Chapter 4 in-book assessment (95% bootstrapped CI [0.10, 1.29]).

Quiz Performance

Near-Transfer Quiz

A Wilcoxon rank-sum test showed no significant difference between the control group ($Mdn = 94$, $n = 106$) and the experimental group ($Mdn = 94$, $n = 103$, $Z = 1.74$, $p = .082$) on the near-transfer quiz. This mean difference was also not significantly different from 0 when controlling for prior performance (bootstrapped 95% CI [−3.09, 0.58]).

We also examined whether the total number of intervention videos watched (across both Chapters 3 and 4) would predict near-transfer performance within the experimental group (Figure 7, lower left; Table 4). We reasoned that even if we did find a significant relationship between videos watched and near-transfer performance, this relationship might be due to unmeasured confounding variables correlated with video watching. For example, high video watchers might be more thorough, hard working, and/or conscientious. For this reason, we controlled for both prior performance and number of nonintervention videos watched later in Chapter 4 (the only chapter that contained nonintervention videos). This allowed us to distinguish the unique effect of watching intervention videos from the effect of video-watching in general.

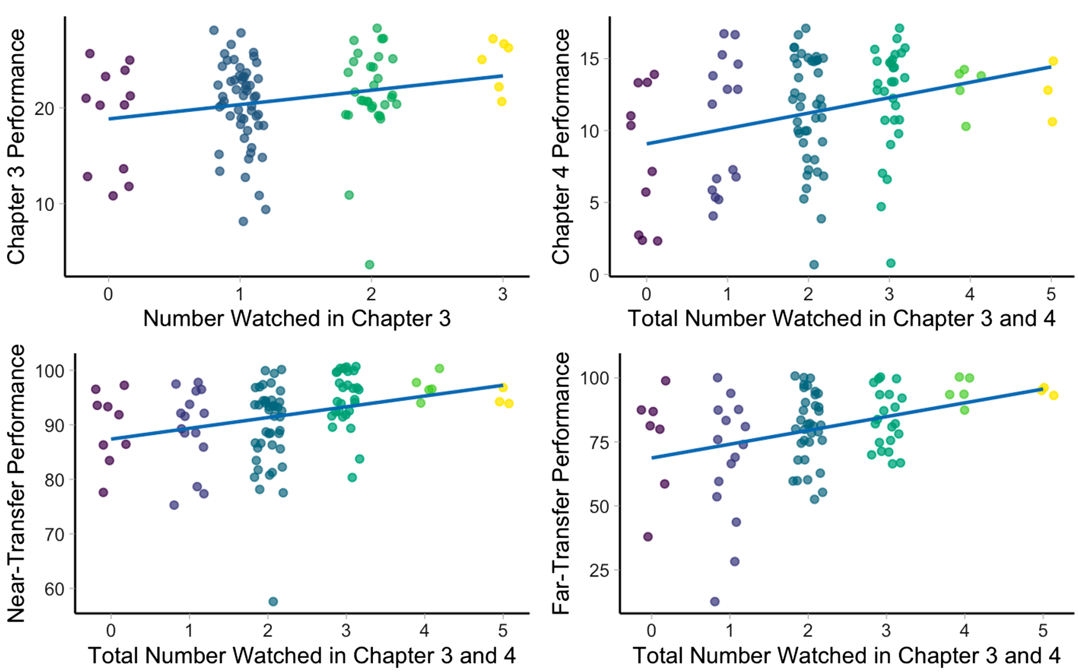
In a bootstrapped analysis controlling for both prior performance and number of nonintervention videos watched, we found that only the total number of intervention videos (95% CI [0.71, 2.71]) significantly predicted students' near-transfer quiz performance. The number of nonintervention videos watched did not significantly predict performance (95% CI [−3.58, 0.93]).

When we conducted a similar analysis for the control group (who obviously did not have intervention videos to watch), the number of nonintervention videos watched also did not significantly predict near-transfer quiz performance (95% CI [−2.21, 2.65]).

Far-Transfer Quiz

A Wilcoxon-rank sum test showed no significant difference between the control group ($Mdn = 81$, $n = 72$) and the experimental group ($Mdn = 81$, $n = 86$. $Z = 0.52$, $p = .602$). This mean difference

Figure 7
Chapters 3 and 4 Near- and Far-Transfer Performance by the Number of Relevant Intervention Videos Watched



Note. See the online article for the color version of this figure.

Table 3

Chapter 3 Performance by the Number of Intervention Videos Watched in Chapter 3 (Experimental Group Only)

Number of intervention videos watched in Chapter 3	Chapter 3 in-book performance			Use of compare and contrast strategy		
	Mdn	*M*	*SD*	*Mdn*	*M*	*SD*
0	20.5	19.17	5.31	3	2.75	1.14
1	21.0	20.35	4.46	4	3.60	0.91
2	21.0	21.45	4.72	4	3.81	1.08
3	25.5	24.67	2.58	5	4.67	0.52

was also not significantly different from 0 when controlling for prior performance (bootstrapped 95% CI [−4.57, 4.97]).

As we did for near-transfer performance, we examined the relationship between total intervention videos watched and students' far-transfer performance, controlling for prior performance and the number of nonintervention videos watched (Figure 8, lower right; Table 4). Consistent with the near-transfer results, the number of intervention videos watched significantly predicted far-transfer performance (95% CI [0.93, 8.06]) but the number of nonintervention videos watched did not (95% CI [−4.32, 6.80]).

When we conducted a similar analysis for the control group, the number of nonintervention videos watched also did not significantly predict far-transfer quiz performance (95% CI [−1.84, 7.28]).

The interscale correlations between Chapter 3 in-book performance, Chapter 4 in-book performance, near-transfer quiz, and far-transfer quiz by condition is shown in Appendix D. The correlations between any two scales ranged from .41 to .68.

Use of Compare and Contrast Strategy

Finally, we investigated whether students' use of compare and contrast strategy differed by condition and number of intervention videos watched. All of these analyses focus on the questions and video-watching behavior from Chapter 3 because that is the chapter that contained open-response questions that prompted the use of comparison strategies. Figure 8 shows the distribution of students' frequency of using compare and contrast strategies by condition. A Wilcoxon rank-sum test found that the experimental group (*Mdn* = 4, $n = 104$) used significantly more compare and contrast strategies than the control group (*Mdn* = 3, $n = 106$; $Z = 6.70$, $p < .001$).

We also examined the relationship between the number of intervention videos watched and use of compare and contrast strategies within the experimental group (Figure 9). The number of intervention videos watched significantly predicted the frequency of using compare and contrast strategies (95% CI [0.24, 0.74]).

Mediation Analyses

We performed a mediation analysis within the experimental group, using the psych package in R (Revelle, 2022), to explore whether students' use of compare and contrast strategies (the strategies modeled in the intervention videos) might mediate the effect of watching the intervention videos on learning (Figure 10). The Chapter 3 in-book assessment was the outcome variable, the number of intervention videos watched in Chapter 3, the predictor, and students' summary scores for use of compare and contrast strategies, the mediator.

The indirect effect of the number of intervention videos watched, through students' use of comparing and contrasting strategies, on in-book assessment was found to be statistically significant (effect = 0.61, 95% CI based on 10,000 bootstrapped estimates = [0.20, 1.15]). That is to say, each additional intervention video students watched resulted in a 0.61 points higher score, on average, on their in-book assessment through the use of comparing and contrasting strategies. The direct effect of the number of intervention videos watched on in-book performance (the c' path) was estimated to be 0.89 points, which was statistically significant, $t(101) = 9.60$, $p < .001$. The total effect (the c path) was estimated to be 1.50, which was also statistically significant, $t(103) = 2.52$, $p = .013$.

On a similar analysis using the near-transfer quiz as the outcome, we found similar results for the indirect effect (Figure 11, Effect = 0.85, 95% CI of 10,000 bootstrapped estimates = [0.09, 1.94]). There was also a significant direct effect, $t(101) = 34.31$, $p < .001$, and total effect, $t(103) = 2.40$, $p = .018$.

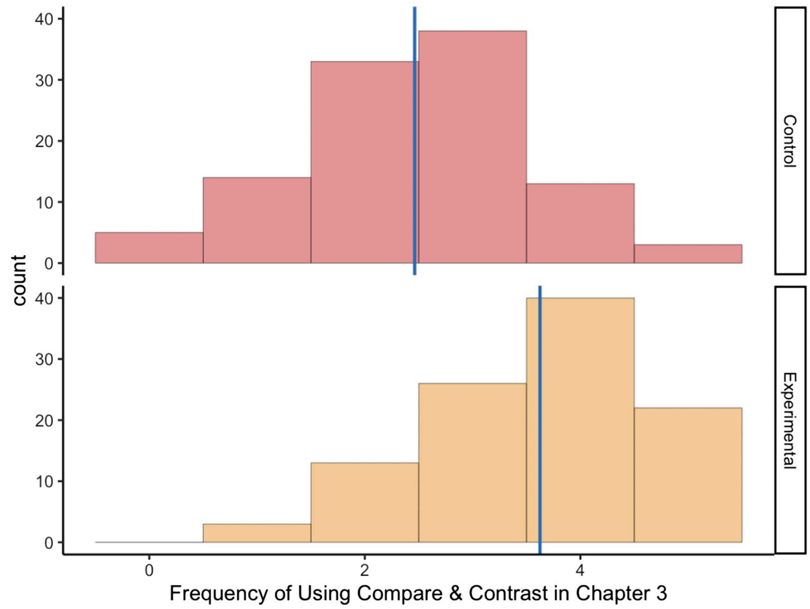
Table 4

Chapter 4 Near- and Far-Transfer Performance by the Total Number of Intervention Videos Watched in Chapters 3 and 4

Total number of intervention videos watched in Chapters 3 and 4	Chapter 4 in-book performance			Near-transfer			Far-transfer		
	Mdn	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*
0	8.5	8.10	4.14	92	89.67	7.70	81	76.43	16.13
1	12.0	10.47	4.72	92	89.47	7.33	75	67.80	24.69
2	12.0	11.43	3.85	92	89.67	7.89	81	80.83	13.61
3	13.0	12.34	3.67	97	94.89	5.07	81	83.43	11.40
4	14.0	13.00	1.73	97	97.00	2.12	94	95.20	5.02
5	13.0	13.00	2.00	94	95.00	1.73	94	95.00	1.73

Figure 8

Students' Use of Compare and Contrast in Chapter 3 by Condition



Note. See the online article for the color version of this figure.

Discussion

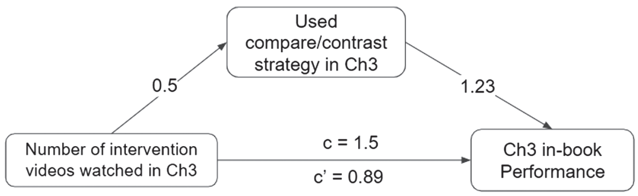
The present study sought to design, integrate and evaluate, in the context of a real course, the effect of a representation-mapping intervention on students' understanding of multiple statistics representations (i.e., texts, histograms, boxplots, R code, word equations). We sought to address three important questions. First, how would we design a mapping intervention that could be integrated into valid learning materials (e.g., an interactive textbook) and thus implemented in a natural class setting over realistic time scales (weeks rather than hours)? Second, would such an intervention facilitate better learning of those specific representations both within the textbook and beyond it (e.g., transfer to class quizzes)? Third, if the mapping intervention did facilitate learning, what could be a possible cognitive mechanism? We will summarize our findings as they relate to the second and third questions, discuss limitations of those findings, and then summarize lessons learned from our implementation and evaluation of a representation-mapping intervention in a realistic learning context.

Figure 9

Students' Use of Compare and Contrast in Chapter 3 by Number of Intervention Videos Watched in Chapter 3

Note. See the online article for the color version of this figure.

Figure 10
Mediation Pathways Using Chapter 3 In-Book Performance as the Outcome Variable



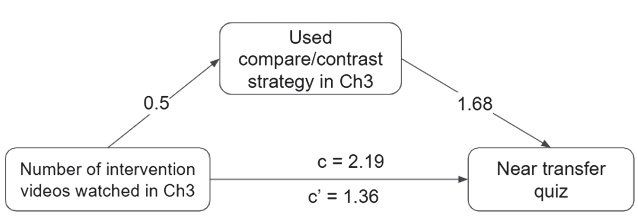
Note. *c* is the path for the total effect of the number of intervention videos watched on performance and *c′* is the path for the direct effect of the number of intervention videos watched in Chapter 3 on Chapter 3 performance.

Overall, the findings showed that representation-mapping intervention videos can facilitate students' learning and transfer in an introductory statistics class—but only if students watch the videos. The difference between the intervention and control condition was only significant for in-book performance (in Chapter 3), but subsequent analyses showed significant relationships between the number of intervention videos watched and all other measures of learning (in-book performance for both chapters, near- and far-transfer). This effect held up even when controlling for prior performance as well as general watching of nonintervention videos in the textbook.

Beyond the simple matter that any intervention must be taken up by the learner for it to be effective, it is also possible that some of the videos were more effective than others. For example, a reliable effect of condition was primarily found for Chapter 3 in-book performance but not for Chapter 4. It is possible that the Chapter 3 intervention videos were more effective due to their content. The content of Chapter 3 intervention videos focused on comparing and mapping across representations of a single distribution (e.g., a single boxplot and histogram). This comparison may have been more fundamental than Chapter 4's content which focused on mapping side-by-side boxplots and faceted histograms.

The fact that the number of intervention videos watched predicted performance after a delay is encouraging. Controlling for students' prior knowledge, the total number of intervention videos students watched, but not the number of nonintervention videos watched, predicted students' performance days and even weeks later on questions that were not directly addressed and explained in the intervention. This rigorous analysis allowed us to conclude that the effect is unlikely to be due simply to "being a better student." It suggests that watching representational-mapping videos can uniquely impact learning. This effect of providing a representational-mapping intervention are in agreement with many findings in the analogy realm, with many laboratory studies demonstrating that alignment and comparison can increase conceptual understanding and transfer because they direct attention to relational structure (e.g., Alfieri et al., 2013; Gentner et al., 2003; Gentner & Maravilla, 2018; Goldwater & Schalk, 2016; Holyoak, 2012; Vendetti et al., 2015).

Figure 11
Mediation Pathways Using the Near-Transfer Quiz Performance as the Outcome Variable



Our exploratory mediation analysis suggests that students' use of compare and contrast strategies may be a mechanism that explains their improved performance. A possible interpretation of these results is that watching the videos led to more frequent use of compare/contrast strategies, and use of these strategies led to better performance. The videos not only modeled explicit representation-mapping strategies but also pointed out the key relational features that emerge when using the strategies.

Our results complement prior results that suggest generalized prompts to "compare and contrast" alone are not particularly effective because novice learners do not spontaneously focus on the most important dimensions of the comparison (Catrambone & Holyoak, 1989). This conclusion aligns with findings in the literature that representational mapping and analogies (Butcher, 2006; Martin et al., 2019; Scheiter & Eitel, 2015) and comparing and contrasting cases (Schwartz et al., 2011) can improve student learning. Our findings provide a specific recommendation to practical instruction: modeling explicit connection making and other guides that ensure students attend to important relational similarities may help students more effectively use comparison strategies later on.

The indirect pathway identified in the mediation analysis suggests that beyond exposing students to such strategies, instructional designers need to think about how to engage students in using the strategies themselves. Even if the connections are obvious to experts and curriculum designers, the connections must be made in the minds of the learners (Fries et al., 2021). Future work should investigate techniques to encourage students' spontaneous application of compare and contrast strategies in authentic learning settings, possibly by asking different types of follow-up questions.

Limitations

Most lab-based research on student learning assumes that the students will watch or read the stimulus materials. The importance of research like ours is that we can clearly see that assumption is flawed. Even knowing that 12%–29% of students do not watch any videos in a given chapter is important for all the instructors out there who are assigning students to watch videos for homework. To motivate students to watch the videos is an important learning engineering challenge that will require experimentation and testing in its own right. We also need to better understand what differs between students who do and do not watch the videos. Learning about students' natural video-watching behavior and its consequences is important.

Although we found that the number of intervention videos watched can predict performance after a long delay, note that students were not given points for watching any videos nor penalized for not watching. This allowed us to capture valid data from video watching. We trust that students were not simply clicking on the videos because they could just have skipped over them. However, the consequence was that we only know part of the effect of these representational alignment videos. This limitation revealed by doing research in valid learning contexts pushes us to consider cognitive and motivational psychology together in future work.

The present study was conducted in a learning environment where a focus on underlying connections was part of the initial design of the textbook. Topping up on an already relationally rich environment is presumably harder, and if these mapping interventions work here, such interventions may be even more effective in other contexts. However, we acknowledge that the effect of representation-mapping interventions may be different in a course without such emphasis, or where students have different expectations (e.g., a focus on memorization or fluency).

Further, although the mapping videos did not cover new content and the book itself already contained many videos, we cannot rule out the simple hypothesis that just having more videos contributes to the differences between the two conditions. For example, one might argue that simply the time spent watching the intervention videos (i.e., more time interacting with learning materials) explains the improvement in performance. However, this explanation seems implausible given that (a) the intervention videos were only 26 min in total length (and few students watched all intervention videos) while students spent approximately 344 min on average working on the two focus chapters; (b) the number of nonintervention videos watched was consistently insignificant and showed a negligible effect size; (c) the mediation analysis suggested comparison as a mechanism for the observed improvement, not simply the presence of more multimedia content. Nonetheless, future studies should include control groups that are given additional videos without the explicit representation-mapping (e.g., providing more examples of boxplots or histograms separately). Alternatively, future studies can also examine whether formats other than videos that encourage representation-mapping could produce similar improvements in learning.

Last, we want to note that our participants were students from a highly selective public university. This population and their learning strategies could be different in important ways. Similarly, although the representation-mapping strategies can be theoretically applied to other domains because they are fundamentally domain-general, we need more research to understand how to translate learning science research for understanding in other domains (e.g., physics, chemistry, writing), learning modalities (e.g., online vs. face to face), and more academically diverse populations.

Lessons Learned From Translating Psychological Insights to Improve Learning in a Real Context

In order to translate a basic finding from cognitive psychology into a learning intervention, we had to find concepts in the domain of statistics that should be helped by representation-mapping. By looking deeply into the discipline, we found ourselves not just using stimuli made for the purpose of examining the psychological construct but broadening the psychological construct to make it useful for teaching statistics.

For example, in tightly controlled lab studies on analogy, there is often one familiar context to start with and then a novel context to which analogies can be drawn. However, in an introductory statistics course, there is no clear base domain to start with (neither histograms nor boxplots can serve as a "base" analogy because both are new to students). Instead, when we used insights from the analogical learning literature to design representation-mapping interventions, we took the basic idea of making relational connections through alignment and comparison and applied that to learning multiple representations.

This application brings together two subcultures within the research community: the multiple representations literature largely based in science, technology, engineering and mathematics education research, with the analogical reasoning literature largely based in cognitive psychology. Our stimuli, focused on representations students have to learn in introductory statistics (e.g., histograms and boxplots), were the product of many converging considerations from both multiple representation pedagogy and analogical reasoning research but also represent a purposeful modification of these research traditions as well.

Conclusion

As educators and learning scientists, our broader purpose is to help students develop deep and transferable understanding in complex domains. There is a long history of research on how hard it is to develop such deep and transferable learning (e.g., Butler et al., 2013; Ericsson, 2006; Paas, 1992). As educators, we want to see improvements in the design of instructional activities that lead to student success. As learning scientists, we want to understand why those improvements lead to better student learning.

The practicing connections framework (Fries et al., 2021) is a pedagogical approach inspired by both of those desires. This framework emphasizes that students should be provided with learning opportunities to practice making the connections that are important to the organizational structure of the domain (Fries et al., 2021). This study demonstrated one way to instantiate this framework, to support making connections between representations through the use of explicit representation-mapping with alignment and comparison. Although representations such as graphs and code are critically important to statistics and data science, the structural connections between them are not obvious to novices. Representation-mapping brings attention to the structural alignment between these representations and shows how the same underlying conceptual schema (e.g., how variation in y is predicted by x) can be seen in multiple forms. This relational understanding can connect the bits of knowledge into a comprehensible framework, allowing novices to start to see coherence in a complex domain.

If we want novices to 1 day become experts, we need to work out how a novice view of the domain develops into the structurally organized expert view (Ericsson et al., 2018). This study served to help novice understanding become more relational in a real class with authentic learning goals. Learning about histograms and boxplots as visualizations of distributions takes a long time but the benefits are also consequential in the long run. As both instructors and researchers we wish all students would simply watch all the videos and do all their homework. But these behaviors are all in service of a cognitive goal: ultimately, we want them to make connections between these concepts in their own minds. Bringing cognitive science principles and rigorous research methodology to a rich learning context, although difficult and complicated, can incrementally move students' knowledge to be more expert-like.

References

Acevedo Nistal, A., Van Dooren, W., Clarebout, G., Elen, J., & Verschaffel, L. (2009). Conceptualising, investigating and stimulating representational flexibility in mathematical problem solving and learning: A critical review. *ZDM Mathematics Education*, *41*(5), 627–636. https://doi.org/10.1007/s11858-009-0189-1

Ackerman, R., & Leiser, D. (2014). The effect of concrete supplements on metacognitive regulation during learning and open-book test taking. *The British Journal of Educational Psychology*, *84*(2), 329–348. https://doi.org/10.1111/bjep.12021

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with MR. *Learning and Instruction*, *16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences*, *11*(1), 25–61. https://doi.org/10.1207/S15327809JLS1101_2

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

Ayres, P. (2020). Something old, something new from cognitive load theory. *Computers in Human Behavior*, *113*, Article 106503. https://doi.org/10.1016/j.chb.2020.106503

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology*, *101*(1), 70–87. https://doi.org/10.1037/a0013247

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction*, *14*(3), 325–341. https://doi.org/10.1016/j.learninstruc.2004.06.006

Bossé, M. J., Adu-Gyamfi, K., & Cheetham, M. R. (2011). Assessing the difficulty of mathematical translations: Synthesizing the literature and novel findings. *International Electronic Journal for Mathematics Education*, *6*(3), 113–133. https://doi.org/10.29333/iejme/264

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, *98*(1), 182–197. https://doi.org/10.1037/0022-0663.98.1.182

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, *105*(2), 290–298. https://doi.org/10.1037/a0031026

Canty, A., & Ripley, B. D. (2021). *boot: Bootstrap R (S-Plus) functions* (Version 1.3-27) [R-package].

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1147–1156. https://doi.org/10.1037/0278-7393.15.6.1147

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/web/packages/pwr/index.html

Chang, B. L., Cromley, J. G., & Tran, N. (2016). Coordinating multiple representations in a reform calculus textbook. *International Journal of Science and Mathematics Education*, *14*(8), 1475–1497. https://doi.org/10.1007/s10763-015-9652-3

Cheng, P. C. H. (2000). 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

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

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

de Jong, T., Ainsworth, S., Dobson, M., van der Hulst, A., Levonen, J., Reimann, P., Sime, J., van Someren, M., Spada, H., & Swaak, J. (1998). Acquiring knowledge in science and mathematics: The use of multiple representations in technology based learning environments. In M. W. van Someren, P. Reimann, & H. P. A. Boshuizen (Eds.), *Learning with multiple representations* (pp. 9–40). Pergamon/Elsevier.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.038

Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (Eds.). (2018). *The Cambridge handbook of expertise and expert performance*. Cambridge University Press. https://doi.org/10.1017/9781316480748

Even, R. (1998). Factors involved in linking representations of functions. *The Journal of Mathematical Behavior*, *17*(1), 105–121. https://doi.org/10.1016/S0732-3123(99)80063-7

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

Gagatsis, A., Elia, I., Panaoura, A., Gravvani, K., & Spyrou, P. (2006). *An empirical four-dimensional model for the understanding of function* [Conference session]. International Group for the Psychology of Mathematics Education.

Gentner, D., & Kurtz, K. J. (2005). Relational categories. In W.-K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin* (pp. 151–175). American Psychological Association. https://doi.org/10.1037/11156-009

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–408. https://doi.org/10.1037/0022-0663.95.2.393

Gentner, D., & Maravilla, F. (2018). Analogical reasoning. In L. J. Ball, & V. A. Thompson (Eds.), *International handbook of thinking and reasoning* (pp. 186–203). Routledge.

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, *142*(7), 729–757. https://doi.org/10.1037/bul0000043

Gray, M. E., & Holyoak, K. J. (2021). Teaching by analogy: From theory to practice. *Mind, Brain and Education: The Official Journal of the International Mind, Brain, and Education Society*, *15*(3), 250–263. https://doi.org/10.1111/mbe.12288

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology*, *24*(2), 95–123. https://doi.org/10.1006/ceps.1998.0987

Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 37–53. https://doi.org/10.1037/a0017683

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, *32*(6), 717–742. https://doi.org/10.1006/jmla.1993.1036

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak, & R. G. Morrison (Eds.), *The oxford handbook of thinking and reasoning, oxford library of psychology* (pp. 234–259). Oxford Academic. https://doi.org/10.1093/oxfordhb/9780199734689.013.0013

Jaeger, A. J., & Wiley, J. (2014). Do illustrations help or harm metacomprehension accuracy? *Learning and Instruction*, *34*, 58–73. https://doi.org/10.1016/j.learninstruc.2014.08.002

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., & Sageman, B. (2013). Finding faults: Analogical comparison supports spatial concept learning in geoscience. *Cognitive Processing*, *14*(2), 175–187. https://doi.org/10.1007/s10339-013-0551-7

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science*, *2*(2), 285–305. https://doi.org/10.1111/j.1756-8765.2009.01053.x

Kozma, R. B., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching*, *34*(9), 949–968. https://doi.org/10.1002/(SICI)1098-2736(199711)34:9<949::AID-TEA7>3.0.CO;2-U

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65–100. https://doi.org/10.1111/j.1551-6708.1987.tb00863.x

Lin, Y. I., Son, J. Y., & Rudd, J. A., II. (2016). Asymmetric translation between multiple representations in chemistry. *International Journal of Science Education*, *38*(4), 644–662. https://doi.org/10.1080/09500693.2016.1144945

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 227–246). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.012

Lowe, R. (2004). Interrogation of a dynamic visualization during learning. *Learning and Instruction*, *14*(3), 257–274. https://doi.org/10.1016/j.learninstruc.2004.06.003

Martin, W., Silander, M., & Rutter, S. (2019). Digital games as sources for science analogies: Learning about energy through play. *Computers & Education*, *130*, 1–12. https://doi.org/10.1016/j.compedu.2018.11.002

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, *60*(1), 95–109. https://doi.org/10.1016/j.compedu.2012.07.011

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

McGee, D. L., & Moore-Russo, D. (2015). Impact of explicit presentation of slopes in three dimensions on students' understanding of derivatives in multivariable calculus. *International Journal of Science and Mathematics Education*, *13*(2), 357–384. https://doi.org/10.1007/s10763-014-9542-0

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Rau, M. A., Michaelis, J. E., & Fay, N. (2015). Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry. *Computers & Education*, *82*, 460–485. https://doi.org/10.1016/j.compedu.2014.12.009

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, *31*(2), 115–121. https://doi.org/10.1207/s15326985ep3102_3

Renkl, A., & Scheiter, K. (2017). Studying visual displays: How to instructionally support learning. *Educational Psychology Review*, *29*(3), 599–621. https://doi.org/10.1007/s10648-015-9340-4

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research* (R package Version 2.2.9). Northwestern University. https://CRAN.R-project.org/package=psych

Richland, L. E. (2015). Linking gestures: Cross-cultural variation during instructional analogies. *Cognition and Instruction*, *33*(4), 295–321. https://doi.org/10.1080/07370008.2015.1091459

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Mathematics. Cognitive supports for analogies in the mathematics classroom. *Science*, *316*(5828), 1128–1129. https://doi.org/10.1126/science.1142103

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, *34*(2), 344–355. https://doi.org/10.3758/BF03193412

Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction*, *36*, 11–26. https://doi.org/10.1016/j.learninstruc.2014.11.002

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). Explaining the modality and contiguity effects: New insights from investigating students' viewing behavior. *Applied Cognitive Psychology*, *24*(2), 226–237. https://doi.org/10.1002/acp.1554

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*(2), 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

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

Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology*, *23*(9), 1227–1243. https://doi.org/10.1002/acp.1526

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, *13*(2), 227–237. https://doi.org/10.1016/S0959-4752(02)00022-1

Seufert, T., & Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Applied Cognitive Psychology*, *20*(3), 321–331. https://doi.org/10.1002/acp.1248

Son, J. Y., Blake, A. B., Fries, L., & Stigler, J. W. (2021). Modeling first: Applying learning science to the teaching of introductory statistics. *Journal of Statistics and Data Science Education*, *29*(1), 4–21. https://doi.org/10.1080/10691898.2020.1844106

Son, J. Y., & Stigler, J. W. (2017–2023). *Statistics and data science: A modeling approach*. https://coursekata.org/

Spiro, R. J. (1988). *Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains*. Technical Report No. 441. Center for the Study of Reading.

Star, J. R., & Rittle-Johnson, B. (2009). Making algebra work: Instructional strategies that deepen student understanding, within and between algebraic representations. *ERS Spectrum*, *27*(2), 11–18. https://dash.harvard.edu/bitstream/handle/1/4889486/ERS%20Resub%20090220.pdf?sequence=1&isAllowed=y

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. https://doi.org/10.1177/016146812012200913

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

van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction*, *16*(3), 199–212. https://doi.org/10.1016/j.learninstruc.2006.03.007

Van Dooren, W., De Bock, D., & Verschaffel, L. (2012). How students understand aspects of linearity. In T. Tso (Ed.), *Proceedings of the 36th conference of the international group for the psychology of mathematics education* (pp. 179–186). Psychology of Mathematics Education.

Van Someren, M. W., Reimann, P., Boshuizen, H., & de Jong, T. (1998). *Learning with Multiple Representations. Advances in learning and instruction series*. Elsevier.

Vendetti, M. S., Matlen, B. J., Richland, L. E., & Bunge, S. A. (2015). Analogical reasoning in the classroom: Insights from cognitive science. *Mind, Brain and Education: The Official Journal of the International Mind, Brain, and Education Society*, *9*(2), 100–106. https://doi.org/10.1111/mbe.12080

Waisman, I., Leikin, M., Shaul, S., & Leikin, R. (2014). Brian activity associated with translation between graphical and symbolic representations of functions in generally gifted and excelling in mathematics adolescents. *International Journal of Science and Mathematics Education*, *12*(3), 669–696. https://doi.org/10.1007/s10763-014-9513-5

Wiley, J. (2019). Picture this! Effects of photographs, diagrams, animations, and sketching on learning and beliefs about learning from a geoscience text. *Applied Cognitive Psychology*, *33*(1), 9–19. https://doi.org/10.1002/acp.3495

Zhang, I. (2023, January 5). *Representational mapping strategies improve learning from an online statistics textbook*. https://osf.io/yxgm9/

(*Appendices follow*)

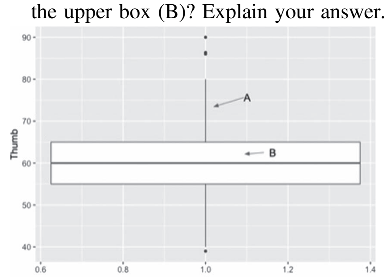
Appendix A

Links to Intervention Videos

Video number	Link to video
Chapter 3 Intervention Video 1	https://vimeo.com/487357715
Chapter 3 Intervention Video 2	https://vimeo.com/487378988
Chapter 3 Intervention Video 3	https://vimeo.com/487384713
Chapter 4 Intervention Video 1	https://vimeo.com/488213872
Chapter 4 Intervention Video 2	https://vimeo.com/488222514

Appendix B

Textbook Questions Examined in the Study

Where in book	Question
Page 3.7	Based on what you've learned, how does a histogram represent variation in the outcome values?
Page 3.7	Based on what you've learned, how does a boxplot represent variation in the outcome values?
Page 3.7	How would you infer the shape of the distribution from a boxplot?
Page 3.7	Looking at the boxplot, where is the data more spread apart (has a greater range): In the upper whisker (A) or in the upper box (B)? Explain your answer. 
Page 3.7	What would the code below do? gf_boxplot(Thumb ~ 100, data = Fingers)
Page 3.7	If I change the code above to gf_boxplot(Thumb ~ "", data = Fingers), what will happen? A. Nothing B. It will return an error C. It will change the value on the *x*-axis D. It will change the value on the *y*-axis
Page 3.8	Compare and contrast bar graphs and histograms. Thinking in particular about the *x*-axis, the *y*-axis, and outcome variables, what features are similar and what are different?
Page 3.8	Notice that a lot of the functions (e.g., gf_histogram, gf_bar, and tally) use the ~. In histograms and bar graphs, what goes after the ~ gets placed on the *x*-axis (the horizontal axis). Is this also true in tally? How so?
Page 3.8	What did you notice about the tallies for categorical versus quantitative variables?
End of Chapter 3 review	Compare and contrast histogram with boxplot. In what ways are they similar and in what ways are they different from each other? (Discuss the strength and weakness of each in terms of describing shape, center, and spread.)
End of Chapter 3 review	If a group of 100 students were given a boxplot and asked to draw a histogram based on the boxplot, what do you think will happen? A. Students will produce 100 identical histograms B. Students' histograms may depict a variety of ranges because the boxplot will not be clear on what the range of the distribution is C. Students' histograms may depict a variety of shapes because the boxplot will not be clear on what the shape of the distribution is D. Students' histograms may depict a variety of ranges and shapes because the boxplot will not be clear on what the range and shape of the distribution is
End of Chapter 3 review	Someone claims that the shape of the distribution of IQ scores is actually bimodal. To verity whether that claim is true, which would be the best visualization to use? A. Boxplot B. Histogram C. Bar graph D. None of the above

(*Appendices continue*)

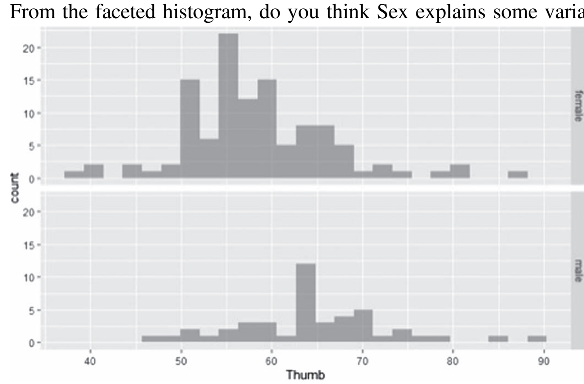
Appendix B (*continued*)

Where in book	Question
End of Chapter 3 review	If you want to know the range for the middle 50% of the data, which visualization would you use? A. Boxplot B. Histogram C. Bar graph D. None of the above
End of Chapter 3 review	If we want to make boxplots that have Thumb on the *y*-axis and Sex on the *x*-axis, which code do you think might work? A. gf_boxplot(Thumb ~ 1, data = Fingers) B. gf_boxplot(1 ~ Thumb, data = Fingers) C. gf_boxplot(Thumb ~ Sex, data = Fingers) D. gf_boxplot(Sex ~ Thumb, data = Fingers)
End of Chapter 3 review	What might be the code for producing two histograms of Thumb in a column, as if the histograms were along the *y*-axis? A. gf_histogram(~ Sex, data = Fingers) %>% gf_facet_grid(.~ Thumb) B. gf_histogram(~ Thumb, data = Fingers) %>% gf_facet_grid(Sex ~ .) C. gf_histogram(~ Sex, data = Fingers) %>% gf_facet_grid(Thumb ~ .) D. None of the above
End of Chapter 3 review	If we told you about a new kind of visualization called a violin plot (the function for it is gf_violin()), how should we write the code to depict Thumb on the *y*-axis? A. gf_violin(Thumb ~ 1, data = Fingers) B. gf_violin(1 ~ Thumb, data = Fingers) C. gf_violin(Y = Thumb, data = Fingers) D. gf_violin(X = Thumb, data = Fingers)
End of Chapter 3 review	If we made a histogram with four bins (bins = 4) would those bins cut the data at the same points as the five number summary?
Page 4.1	Can we see how thumb lengths vary by sex in the histogram we made? Why not?
Page 4.1	Although better than the single histogram, it is still not easy to compare thumb length across the two side-by-side histograms. We have to keep numbers in our mind as we look back and forth. Would it be more helpful if the histograms were stacked vertically (one above the other)? Why?
Page 4.1	Take a look at the density histograms you made. Do thumb lengths vary by sex? In what way? Is there still variation in thumb length among people of the same sex?
Page 4.1	What features of the faceted histograms do you look at to judge whether one variable explains variation in another variable?
Page 4.2	You previously learned that categorical variables should go into gf_facet_grid(). Now you know that Sex is not only categorical, it is also the explanatory variable. Why is it useful to split up the histograms by the explanatory variable?
Page 4.2	Actually, you can have categorical outcome variables! It's just that they would not be represented in a histogram. What kind of visualization (e.g., graph or plot) would be most appropriate for examining a categorical outcome variable? Why?
Page 4.4	In the plot above in which the male box is higher than the female box, how should we interpret the placement of the boxes? How should we interpret this visual feature?
Page 4.4	If there was no difference between groups, how would you expect the boxes to be positioned in a boxplot?
Page 4.4	What do you think the boxplot for Thumb length by Job might look like? Job is a categorical variable with three levels (no job, part-time, and full-time). How would the boxplot look different from the one for Sex, which had just two levels?
Page 4.4	Why is the full-time box so different from the other two?
Page 4.5	Based on the example interpretation for thumb length ("variation in Thumb length is explained by variation in Sex plus variation in other stuff"), try writing an interpretation for this word equation: Happiness in countries = Health of individuals + other stuff
Page 4.5	Happiness in countries = Wealth of country + Environmental beauty + other stuff
Page 4.5	Try writing a word equation for health of housekeepers. What are some explanatory variables that might explain variation in health?

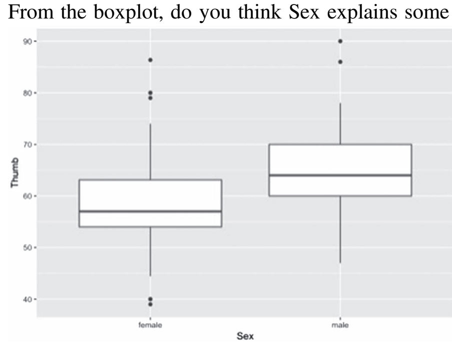
(*Appendices continue*)

Appendix B (*continued*)

Where in book	Question
Where students saw the Chapter 4 intervention videos (the start of Chapter 4 in book assessment)	
Page 4.5	From the faceted histogram, do you think Sex explains some variation in Thumb? Why or why not?



Where in book	Question
Page 4.5	From the boxplot, do you think Sex explains some variation in Thumb? Why or why not?



Assume you know the cutoffs for outliers in a data set.

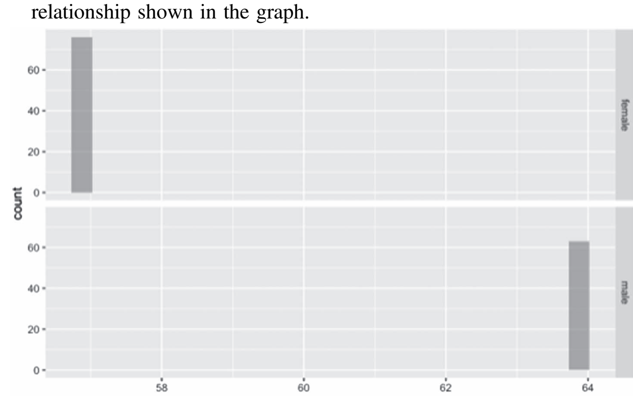
If you want to find the exact number of outliers in your data set, which graph(s) should you use?

A. Histogram only
B. Boxplot only
C. Both histogram and boxplot
D. Either histogram or boxplot

If you want to find the exact numerical value of outliers in your data set, which graph(s) should you use?

A. Histogram only
B. Boxplot only
C. Both histogram and boxplot
D. Either histogram or boxplot

Where in book	Question
Page 4.5	Given the graph below that shows the faceted histogram of Thumb by Sex, write a word equation to express the relationship shown in the graph.



Where in book	Question
Page 4.8	Do you think we could explain some of the variation in thumb length with variation in height? Could we use the same approach with Height as we used with Sex? Why or why not?
Page 4.8	If we created a faceted grid of histograms of Thumb based on Height2Group, what would you expect to see?
Page 4.8	Which Height variable (the three-category variable or the two-category variable) explains more variation in Thumb length? How can you tell? (they always chose the answer in multiple choice above)

(*Appendices continue*)

Appendix B (*continued*)

Where in book	Question
End of Chapter 4 review	Originally, these data were collected by a group of researchers interested in sleep but let's say a researcher comes upon this data and is interested in using it to explain variation in students' levels of happiness. Why might someone want to explain variation in happiness? What other variables in this data frame might be meaningfully related to happiness?
End of Chapter 4 review	Below are two sets of faceted histograms. On the left is Happiness faceted by GPA3Group and on the right is Happiness faceted by Stress. Which variable seems better at explaining variation in Happiness: GPA3Group or Stress? Why do you think that variable is a better explanatory variable? (Cite features of the histograms to support your answer.)
End of Chapter 4 review	A veterinarian measures the resting heart rate of all of his dog clientele and records the breeds of the dogs. He categorizes each dogs as belonging to a "small breed" or "large breed." He organizes all of this information in a data frame called Dogs. To visualize his data, he created a faceted histogram using this code: gf_histogram(~ Heartrate, data = Dogs) %>% gf_facet_grid(Breedsize ~ .) What kind of variable is Heartrate? A. Quantitative B. Qualitative C. Categorical D. Discrete E. You can't tell based on the information given
End of Chapter 4 review	What will appear on the *x*-axis in the graph made by the veterinarian? A. Heart rates of dogs from the Dogs data set B. Breed sizes of dogs from the Dogs data set C. Dogs D. gf_histogram E. Frequency of different values of heart rate
End of Chapter 4 review	How will the faceted histograms appear? A. The two histograms will be side by side along the *x*-axis B. The two histograms will be stacked above and below along the *y*-axis C. It will be completely random D. There is no way to tell unless I run the code
End of Chapter 4 review	If the veterinarian wanted to compare the heart rates of different breed sizes by putting two histograms side by side on the *x*-axis, how should he change the code for the faceted histogram above? A. Do not change the code; it already does this B. Move Heartrate before the tilde in the first line of code: gf_histogram(Heartrate ~ ., data = Dogs) C. Move breedsize after the tilde in the second line of code: gf_facet_grid(. ~ breedSize) D. Histograms cannot be put side by side because it is a frequency distribution
End of Chapter 4 review	Which of the following graphs shows outcome values that are uniformly distributed across the entire range? A. Histogram with equal heights across all bins B. Boxplot with equal-length whiskers and sections of box C. Histogram with different height bins D. Boxplot with whiskers of different lengths
End of Chapter 4 review	A powerlifting coach records how much each of his athletes can bench press in a data set and includes their sex. He wants to find out if sex can explain any variation in an athlete's bench press. Which graph can he use? Select all that apply. A. One histogram B. One boxplot C. Faceted histogram by sex D. Side-by-side boxplots by sex E. Faceted histogram by bench press F. Side-by-side boxplots by bench press
End of Chapter 4 review	Based on his data visualizations, the coach discovers that Sex seems to explain some variation in how much his athletes can bench press. Which of the following data visualizations show that Sex explains variation in bench press? Select all that apply. A. a visualization that shows women bench pressing more than men B. a visualization that shows men and women bench pressing similar amounts C. a visualization that shows that the amount the athletes can bench press varies widely D. a visualization that shows men bench pressing more than women E. a visualization that shows that the amount the athletes can bench press does not vary much
End of Chapter 4 review	How would you tell by looking at the graphs if breed size explained variation in resting heart rate? What would you look for in the graphs?
End of Chapter 4 review	Suppose the heart rate of the dog can be better predicted if we knew the breed size of the dog, the dog's physical health condition, and other stuff. Write the word equation to express this.

(*Appendices continue*)

Appendix C

Results of Nonnormality Test of Performance Measures

	Shapiro–Wilk test	
Performance measure	*W*	*p*
Chapters 3 and 4 page view	0.93	<.001
Chapter 2 in-book performance	0.79	<.001
Chapter 3 in-book performance	0.95	<.001
Chapter 4 in-book performance	0.92	<.001
Near transfer	0.86	<.001
Far transfer	0.91	<.001
Use of compare and contrast	0.93	<.001

Appendix D

Statistics for Interscale Correlations

Table D1

Interscale Correlations for the Control Group

Assessment	Chapter 3	Chapter 4	Near transfer	Far transfer
Chapter 3	—			
Chapter 4	0.64***	—		
Near transfer	0.46***	0.41***	—	
Far transfer	0.47***	0.44***	0.46***	—

*** $p < .001$ (two-tailed).

Table D2

Interscale Correlations for the Experimental Group

Assessment	Chapter 3	Chapter 4	Near transfer	Far transfer
Chapter 3	—			
Chapter 4	0.68***	—		
Near transfer	0.56***	0.41***	—	
Far transfer	0.49***	0.44***	0.48***	—

*** $p < .001$ (two-tailed).

Appendix E

Complete Regression Results

Table E1

Regression Results Using Chapter 3 Performance as the Outcome

Predictor	*b*	*b* 95% CI [*LL*, *UL*]	Fit
(Intercept)	4.19**	[−2.05, 8.77]	
Condition	1.57**	[0.50, 2.67]	
Preintervention performance	0.56**	[0.40, 0.78]	
			$R^2 = .338$** 95% CI [.23, .48]

Note. A significant *b*-weight indicates the semipartial correlation is also significant. *b* represents unstandardized regression weights. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

(*Appendices continue*)

Table E2
Regression Results Using Chapter 3 Performance as the Outcome (Within Experimental)

Predictor	b	b 95% CI [LL, UL]	Fit
(Intercept)	4.67*	[−2.40, 9.67]	
Number of Chapter 3 intervention videos watched	1.09*	[0.12, 1.98]	
Preintervention performance	0.54**	[0.36, 0.80]	
			R^2 = .336** 95% CI [.22, .49]

Note. b represents unstandardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table E3
Regression Results Using Chapter 4 Performance as the Outcome

Predictor	b	b 95% CI [LL, UL]	Fit
(Intercept)	1.87	[−3.06, 5.87]	
Condition	−0.01	[−0.98, 0.98]	
Preintervention performance	0.35**	[0.21, 0.53]	
			R^2 = .181** 95% CI [.08, .33]

Note. b represents unstandardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table E4
Regression Results Using Chapter 4 Performance as the Outcome (Within Experimental)

Predictor	b	b 95% CI [LL, UL]	Fit
(Intercept)	−2.43	[−6.59, 0.74]	
Total number of intervention videos watched	0.69**	[0.10, 1.29]	
Preintervention performance	0.46**	[0.33, 0.62]	
			R^2 = .339** 95% CI [.20, .49]

Note. b represents unstandardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table E5
Regression Results Using Near-Transfer Quiz Performance as the Outcome

Predictor	b	b 95% CI [LL, UL]	Fit
(Intercept)	75.36**	[65.11, 82.91]	
Condition	−1.26	[−3.09, 0.58]	
Preintervention performance	0.65**	[0.38, 1.01]	
			R^2 = .185** 95% CI [.09, .33]

Note. b represents unstandardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

(*Appendices continue*)

Table E6
Regression Results Using Near-Transfer Quiz Performance as the Outcome (Within Experimental)

Predictor	b	b 95% CI [LL, UL]	Fit
(Intercept)	69.92**	[56.48, 81.01]	
Total number of intervention videos watched	1.63*	[0.71, 2.71]	
Number of nonintervention videos watched	−1.19	[−3.58, 0.93]	
Preintervention performance	0.71**	[0.29, 1.22]	
			R^2 = .294** 95% CI [.14, .51]

Note. b represents unstandardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table E7
Regression Results Using Near-Transfer Quiz Performance as the Outcome (Within Control)

Predictor	b	b 95% CI [LL, UL]	Fit
(Intercept)	77.61**	[63.68, 86.02]	
Number of nonintervention videos watched	0.27	[−2.21, 2.65]	
Preintervention performance	0.56**	[0.24, 1.08]	
			R^2 = .139** 95% CI [.05, .33]

Note. b represents unstandardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table E8
Regression Results Using Far-Transfer Quiz Performance as the Outcome

Predictor	b	b 95% CI [LL, UL]	Fit
(Intercept)	49.56**	[21.15, 69.56]	
Condition	0.13	[−4.57, 4.97]	
Preintervention performance	1.13**	[0.41, 2.15]	
			R^2 = .113** 95% CI [.02, .31]

Note. b represents unstandardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table E9
Regression Results Using Far-Transfer Quiz Performance as the Outcome (Within Experimental)

Predictor	b	b 95% CI [LL, UL]	Fit
(Intercept)	45.98**	[18.01, 72.14]	
Total number of intervention videos watched	4.15*	[0.93, 8.06]	
Number of nonintervention videos watched	1.85	[−4.32, 6.80]	
Preintervention performance	0.88**	[−0.11, 1.98]	
			R^2 = .187** 95% CI [.07, .43]

Note. b represents unstandardized regression weights. CI = confidence interval; LL = lower limit; UL = upper limit.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

(*Appendices continue*)

Table E10
Regression Results Using Far-Transfer Quiz Performance as the Outcome (Within Control)

Predictor	*b*	*b* 95% CI [*LL*, *UL*]	Fit
(Intercept)	49.06**	[−1.21, 72.73]	
Number of nonintervention videos watched	2.68	[−1.84, 7.28]	
Preintervention performance	1.06**	[0.22, 2.88]	
			$R^2 = .148^{**}$ 95% CI [.03, .45]

Note. *b* represents unstandardized regression weights. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Received August 1, 2022
Revision received February 7, 2023
Accepted February 14, 2023 ■